

COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
AND DIAGNOSIS OF TUBERCULOSIS

5 CROSS-REFERENCE TO RELATED APPLICATIONS

906 7  
A1

10084843 " 022502  
205220 " 24848001

This application is a continuation-in-part of U.S. Application No. 9/025,197, filed February 18, 1998; which is a continuation-in-part of U.S. Application No. 08/942,578, filed October 1, 1997; which is a continuation-in-part of U.S. Application No. 08/818,112, filed March 13, 1997; which is a continuation-in-part of U.S. Application No. 08/730,510, filed October 11, 1996; which claims priority from PCT Application No. PCT/US 96/14674, filed August 30, 1996; and is a continuation-in-part of U.S. Application No. 08/680,574, filed July 12, 1996; which is a continuation-in-part of U.S. Application No. 08/659,683, filed June 5, 1996; which is a continuation-in-part of U.S. Application No. 08/620,874, filed March 22, 1996, now abandoned; which is a continuation-in-part of U.S. Application No. 08/533,634, filed September 22, 1995, now abandoned; which is a continuation-in-part of U.S. Application No. 08/523,436, filed September 1, 1995, now abandoned.

TECHNICAL FIELD

20 The present invention relates generally to detecting, treating and preventing *Mycobacterium tuberculosis* infection. The invention is more particularly related to polypeptides comprising a *Mycobacterium tuberculosis* antigen, or a portion or other variant thereof, and the use of such polypeptides for diagnosing and vaccinating against *Mycobacterium tuberculosis* infection.

25

BACKGROUND OF THE INVENTION

Tuberculosis is a chronic, infectious disease, that is generally caused by infection with *Mycobacterium tuberculosis*. It is a major disease in developing countries, as well as an increasing problem in developed areas of the world, with about 30 8 million new cases and 3 million deaths each year. Although the infection may be asymptomatic for a considerable period of time, the disease is most commonly

manifested as an acute inflammation of the lungs, resulting in fever and a nonproductive cough. If left untreated, serious complications and death typically result.

Although tuberculosis can generally be controlled using extended antibiotic therapy, such treatment is not sufficient to prevent the spread of the disease. 5 Infected individuals may be asymptomatic, but contagious, for some time. In addition, although compliance with the treatment regimen is critical, patient behavior is difficult to monitor. Some patients do not complete the course of treatment, which can lead to ineffective treatment and the development of drug resistance.

Inhibiting the spread of tuberculosis requires effective vaccination and 10 accurate, early diagnosis of the disease. Currently, vaccination with live bacteria is the most efficient method for inducing protective immunity. The most common Mycobacterium employed for this purpose is *Bacillus Calmette-Guerin* (BCG), an avirulent strain of *Mycobacterium bovis*. However, the safety and efficacy of BCG is a source of controversy and some countries, such as the United States, do not vaccinate 15 the general public. Diagnosis is commonly achieved using a skin test, which involves intradermal exposure to tuberculin PPD (protein-purified derivative). Antigen-specific T cell responses result in measurable induration at the injection site by 48-72 hours after injection, which indicates exposure to Mycobacterial antigens. Sensitivity and specificity have, however, been a problem with this test, and individuals vaccinated 20 with BCG cannot be distinguished from infected individuals.

While macrophages have been shown to act as the principal effectors of *M. tuberculosis* immunity, T cells are the predominant inducers of such immunity. The essential role of T cells in protection against *M. tuberculosis* infection is illustrated by the frequent occurrence of *M. tuberculosis* in AIDS patients, due to the depletion of 25 CD4 T cells associated with human immunodeficiency virus (HIV) infection. Mycobacterium-reactive CD4 T cells have been shown to be potent producers of gamma-interferon (IFN- $\gamma$ ), which, in turn, has been shown to trigger the anti-mycobacterial effects of macrophages in mice. While the role of IFN- $\gamma$  in humans is less clear, studies have shown that 1,25-dihydroxy-vitamin D3, either alone or in 30 combination with IFN- $\gamma$  or tumor necrosis factor-alpha, activates human macrophages

10084843-022502

to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN- $\gamma$  stimulates human macrophages to make 1,25-dihydroxy-vitamin D3. Similarly, IL-12 has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection see Chan and Kaufmann in  
 5 *Tuberculosis: Pathogenesis, Protection and Control*, Bloom (ed.), ASM Press, Washington, DC, 1994.

Accordingly, there is a need in the art for improved vaccines and methods for preventing, treating and detecting tuberculosis. The present invention fulfills these needs and further provides other related advantages.

10

#### SUMMARY OF THE INVENTION

Briefly stated, this invention provides compounds and methods for preventing and diagnosing tuberculosis. In one aspect, polypeptides are provided comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of  
 15 such an antigen that differs only in conservative substitutions and/or modifications. In one embodiment of this aspect, the soluble antigen has one of the following N-terminal sequences:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 120)
- 20 (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 123)
- 25 (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 124)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 125)
- 30 (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser; (SEQ ID No. 126)

10084843-022502

- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn; (SEQ ID No. 128)
- (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 134)
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) or
- (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

wherein Xaa may be any amino acid.

In a related aspect, polypeptides are provided comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, the antigen having one of the following N-terminal sequences:

- (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137) or
- (n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID No. 129)

wherein Xaa may be any amino acid.

In another embodiment, the soluble *M. tuberculosis* antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101 or a complement thereof under moderately stringent conditions.

In a related aspect, the polypeptides comprise an immunogenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, wherein the antigen comprises an

10084843.022502

amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 339, 333, 335, 337, 339 and 341, the complements of said sequences, and DNA sequences that hybridize to a sequence  
5 recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341 or a complement thereof under moderately stringent conditions.

In related aspects, DNA sequences encoding the above polypeptides, expression vectors comprising these DNA sequences and host cells transformed or  
10 transfected with such expression vectors are also provided.

In another aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, an inventive polypeptide and a known *M. tuberculosis* antigen.

Within other aspects, the present invention provides pharmaceutical  
15 compositions that comprise one or more of the above polypeptides, or a DNA molecule encoding such polypeptides, and a physiologically acceptable carrier. The invention also provides vaccines comprising one or more of the polypeptides as described above and a non-specific immune response enhancer, together with vaccines comprising one or more DNA sequences encoding such polypeptides and a non-specific immune  
20 response enhancer.

In yet another aspect, methods are provided for inducing protective immunity in a patient, comprising administering to a patient an effective amount of one or more of the above polypeptides.

In further aspects of this invention, methods and diagnostic kits are  
25 provided for detecting tuberculosis in a patient. The methods comprise contacting dermal cells of a patient with one or more of the above polypeptides and detecting an immune response on the patient's skin. The diagnostic kits comprise one or more of the above polypeptides in combination with an apparatus sufficient to contact the polypeptide with the dermal cells of a patient.

10084843-022502

In yet other aspects, methods are provided for detecting tuberculosis in a patient, such methods comprising contacting dermal cells of a patient with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and detecting an immune response on the patient's skin. Diagnostic kits for use in such methods are also provided.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

#### BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE IDENTIFIERS

Figure 1A and B illustrate the stimulation of proliferation and interferon- $\gamma$  production in T cells derived from a first and a second *M. tuberculosis*-immune donor, respectively, by the 14 Kd, 20 Kd and 26 Kd antigens described in Example 1.

Figure 2 illustrates the stimulation of proliferation and interferon- $\gamma$  production in T cells derived from an *M. tuberculosis*-immune individual by the two representative polypeptides TbRa3 and TbRa9.

Figures 3A-D illustrate the reactivity of antisera raised against secretory *M. tuberculosis* proteins, the known *M. tuberculosis* antigen 85b and the inventive antigens Tb38-1 and TbH-9, respectively, with *M. tuberculosis* lysate (lane 2), *M. tuberculosis* secretory proteins (lane 3), recombinant Tb38-1 (lane 4), recombinant TbH-9 (lane 5) and recombinant 85b (lane 5).

Figure 4A illustrates the stimulation of proliferation in a TbH-9-specific T cell clone by secretory *M. tuberculosis* proteins, recombinant TbH-9 and a control antigen, TbRa11.

10084843 "022502

Figure 4B illustrates the stimulation of interferon- $\gamma$  production in a TbH-9-specific T cell clone by secretory *M. tuberculosis* proteins, PPD and recombinant TbH-9.

Figures 5A and B illustrate the stimulation of proliferation and interferon- $\gamma$  production in TbH9-specific T cells by the fusion protein TbH9-Tb38-1.

Figures 6A and B illustrate the stimulation of proliferation and interferon- $\gamma$  production in Tb38-1-specific T cells by the fusion protein TbH9-Tb38-1.

Figures 7A and B illustrate the stimulation of proliferation and interferon- $\gamma$  production in T cells previously shown to respond to both TbH-9 and Tb38-1 by the fusion protein TbH9-Tb38-1.

Figures 8A and B illustrate the stimulation of proliferation and interferon- $\gamma$  production in T cells derived from a first *M. tuberculosis*-immune individual by the representative polypeptides XP-1, RDIF6, RDIF8, RDIF10 and RDIF11.

Figures 9A and B illustrate the stimulation of proliferation and interferon- $\gamma$  production in T cells derived from a second *M. tuberculosis*-immune individual by the representative polypeptides XP-1, RDIF6, RDIF8, RDIF10 and RDIF11.

SEQ. ID NO. 1 is the DNA sequence of TbRa1.

SEQ. ID NO. 2 is the DNA sequence of TbRa10.

SEQ. ID NO. 3 is the DNA sequence of TbRa11.

SEQ. ID NO. 4 is the DNA sequence of TbRa12.

SEQ. ID NO. 5 is the DNA sequence of TbRa13.

SEQ. ID NO. 6 is the DNA sequence of TbRa16.

SEQ. ID NO. 7 is the DNA sequence of TbRa17.

SEQ. ID NO. 8 is the DNA sequence of TbRa18.

SEQ. ID NO. 9 is the DNA sequence of TbRa19.

SEQ. ID NO. 10 is the DNA sequence of TbRa24.

SEQ. ID NO. 11 is the DNA sequence of TbRa26.

SEQ. ID NO. 12 is the DNA sequence of TbRa28.

205220" E484800T

SEQ. ID NO. 13 is the DNA sequence of TbRa29.

SEQ. ID NO. 14 is the DNA sequence of TbRa2A.

SEQ. ID NO. 15 is the DNA sequence of TbRa3.

SEQ. ID NO. 16 is the DNA sequence of TbRa32.

5 SEQ. ID NO. 17 is the DNA sequence of TbRa35.

SEQ. ID NO. 18 is the DNA sequence of TbRa36.

SEQ. ID NO. 19 is the DNA sequence of TbRa4.

SEQ. ID NO. 20 is the DNA sequence of TbRa9.

SEQ. ID NO. 21 is the DNA sequence of TbRaB.

10 SEQ. ID NO. 22 is the DNA sequence of TbRaC.

SEQ. ID NO. 23 is the DNA sequence of TbRaD.

SEQ. ID NO. 24 is the DNA sequence of YYWCPG.

SEQ. ID NO. 25 is the DNA sequence of AAMK.

SEQ. ID NO. 26 is the DNA sequence of TbL-23.

15 SEQ. ID NO. 27 is the DNA sequence of TbL-24.

SEQ. ID NO. 28 is the DNA sequence of TbL-25.

SEQ. ID NO. 29 is the DNA sequence of TbL-28.

SEQ. ID NO. 30 is the DNA sequence of TbL-29.

SEQ. ID NO. 31 is the DNA sequence of TbH-5.

20 SEQ. ID NO. 32 is the DNA sequence of TbH-8.

SEQ. ID NO. 33 is the DNA sequence of TbH-9.

SEQ. ID NO. 34 is the DNA sequence of TbM-1.

SEQ. ID NO. 35 is the DNA sequence of TbM-3.

SEQ. ID NO. 36 is the DNA sequence of TbM-6.

25 SEQ. ID NO. 37 is the DNA sequence of TbM-7.

SEQ. ID NO. 38 is the DNA sequence of TbM-9.

SEQ. ID NO. 39 is the DNA sequence of TbM-12.

SEQ. ID NO. 40 is the DNA sequence of TbM-13.

SEQ. ID NO. 41 is the DNA sequence of TbM-14.

30 SEQ. ID NO. 42 is the DNA sequence of TbM-15.

205220 "E4B4B00F



SEQ. ID NO. 43 is the DNA sequence of TbH-4.

SEQ. ID NO. 44 is the DNA sequence of TbH-4-FWD.

SEQ. ID NO. 45 is the DNA sequence of TbH-12.

SEQ. ID NO. 46 is the DNA sequence of Tb38-1.

5 SEQ. ID NO. 47 is the DNA sequence of Tb38-4.

SEQ. ID NO. 48 is the DNA sequence of TbL-17.

SEQ. ID NO. 49 is the DNA sequence of TbL-20.

SEQ. ID NO. 50 is the DNA sequence of TbL-21.

SEQ. ID NO. 51 is the DNA sequence of TbH-16.

10 SEQ. ID NO. 52 is the DNA sequence of DPEP.

SEQ. ID NO. 53 is the deduced amino acid sequence of DPEP.

SEQ. ID NO. 54 is the protein sequence of DPV N-terminal Antigen.

SEQ. ID NO. 55 is the protein sequence of AVGS N-terminal Antigen.

SEQ. ID NO. 56 is the protein sequence of AAMK N-terminal Antigen.

15 SEQ. ID NO. 57 is the protein sequence of YYWC N-terminal Antigen.

SEQ. ID NO. 58 is the protein sequence of DIGS N-terminal Antigen.

SEQ. ID NO. 59 is the protein sequence of AEES N-terminal Antigen.

SEQ. ID NO. 60 is the protein sequence of DPEP N-terminal Antigen.

SEQ. ID NO. 61 is the protein sequence of APKT N-terminal Antigen.

20 SEQ. ID NO. 62 is the protein sequence of DPAS N-terminal Antigen.

SEQ. ID NO. 63 is the deduced amino acid sequence of TbRa1.

SEQ. ID NO. 64 is the deduced amino acid sequence of TbRa10.

SEQ. ID NO. 65 is the deduced amino acid sequence of TbRa11.

SEQ. ID NO. 66 is the deduced amino acid sequence of TbRa12.

25 SEQ. ID NO. 67 is the deduced amino acid sequence of TbRa13.

SEQ. ID NO. 68 is the deduced amino acid sequence of TbRa16.

SEQ. ID NO. 69 is the deduced amino acid sequence of TbRa17.

SEQ. ID NO. 70 is the deduced amino acid sequence of TbRa18.

SEQ. ID NO. 71 is the deduced amino acid sequence of TbRa19.

30 SEQ. ID NO. 72 is the deduced amino acid sequence of TbRa24.

205220 "CH818001

SEQ. ID NO. 73 is the deduced amino acid sequence of TbRa26.  
SEQ. ID NO. 74 is the deduced amino acid sequence of TbRa28.  
SEQ. ID NO. 75 is the deduced amino acid sequence of TbRa29.  
SEQ. ID NO. 76 is the deduced amino acid sequence of TbRa2A.  
5 SEQ. ID NO. 77 is the deduced amino acid sequence of TbRa3.  
SEQ. ID NO. 78 is the deduced amino acid sequence of TbRa32.  
SEQ. ID NO. 79 is the deduced amino acid sequence of TbRa35.  
SEQ. ID NO. 80 is the deduced amino acid sequence of TbRa36.  
SEQ. ID NO. 81 is the deduced amino acid sequence of TbRa4.  
10 SEQ. ID NO. 82 is the deduced amino acid sequence of TbRa9.  
SEQ. ID NO. 83 is the deduced amino acid sequence of TbRaB.  
SEQ. ID NO. 84 is the deduced amino acid sequence of TbRaC.  
SEQ. ID NO. 85 is the deduced amino acid sequence of TbRaD.  
SEQ. ID NO. 86 is the deduced amino acid sequence of YYWCPG.  
15 SEQ. ID NO. 87 is the deduced amino acid sequence of TbAAMK.  
SEQ. ID NO. 88 is the deduced amino acid sequence of Tb38-1.  
SEQ. ID NO. 89 is the deduced amino acid sequence of TbH-4.  
SEQ. ID NO. 90 is the deduced amino acid sequence of TbH-8.  
SEQ. ID NO. 91 is the deduced amino acid sequence of TbH-9.  
20 SEQ. ID NO. 92 is the deduced amino acid sequence of TbH-12.  
SEQ. ID NO. 93 is the amino acid sequence of Tb38-1 Peptide 1.  
SEQ. ID NO. 94 is the amino acid sequence of Tb38-1 Peptide 2.  
SEQ. ID NO. 95 is the amino acid sequence of Tb38-1 Peptide 3.  
SEQ. ID NO. 96 is the amino acid sequence of Tb38-1 Peptide 4.  
25 SEQ. ID NO. 97 is the amino acid sequence of Tb38-1 Peptide 5.  
SEQ. ID NO. 98 is the amino acid sequence of Tb38-1 Peptide 6.  
SEQ. ID NO. 99 is the DNA sequence of DPAS.  
SEQ. ID NO. 100 is the deduced amino acid sequence of DPAS.  
SEQ. ID NO. 101 is the DNA sequence of DPV.  
30 SEQ. ID NO. 102 is the deduced amino acid sequence of DPV.

10084843 022502 205220 E1818001

SEQ. ID NO. 103 is the DNA sequence of ESAT-6.

SEQ. ID NO. 104 is the deduced amino acid sequence of ESAT-6.

SEQ. ID NO. 105 is the DNA sequence of TbH-8-2.

SEQ. ID NO. 106 is the DNA sequence of TbH-9FL.

5 SEQ. ID NO. 107 is the deduced amino acid sequence of TbH-9FL.

SEQ. ID NO. 108 is the DNA sequence of TbH-9-1.

SEQ. ID NO. 109 is the deduced amino acid sequence of TbH-9-1.

SEQ. ID NO. 110 is the DNA sequence of TbH-9-4.

SEQ. ID NO. 111 is the deduced amino acid sequence of TbH-9-4.

10 SEQ. ID NO. 112 is the DNA sequence of Tb38-1F2 IN.

SEQ. ID NO. 113 is the DNA sequence of Tb38-2F2 RP.

SEQ. ID NO. 114 is the deduced amino acid sequence of Tb37-FL.

SEQ. ID NO. 115 is the deduced amino acid sequence of Tb38-IN.

SEQ. ID NO. 116 is the DNA sequence of Tb38-1F3.

15 SEQ. ID NO. 117 is the deduced amino acid sequence of Tb38-1F3.

SEQ. ID NO. 118 is the DNA sequence of Tb38-1F5.

SEQ. ID NO. 119 is the DNA sequence of Tb38-1F6.

SEQ. ID NO. 120 is the deduced N-terminal amino acid sequence of DPV.

SEQ. ID NO. 121 is the deduced N-terminal amino acid sequence of AVGS.

20 SEQ. ID NO. 122 is the deduced N-terminal amino acid sequence of AAMK.

SEQ. ID NO. 123 is the deduced N-terminal amino acid sequence of YYWC.

SEQ. ID NO. 124 is the deduced N-terminal amino acid sequence of DIGS.

SEQ. ID NO. 125 is the deduced N-terminal amino acid sequence of AEES.

SEQ. ID NO. 126 is the deduced N-terminal amino acid sequence of DPEP.

25 SEQ. ID NO. 127 is the deduced N-terminal amino acid sequence of APKT.

SEQ. ID NO. 128 is the deduced amino acid sequence of DPAS.

SEQ. ID NO. 129 is the protein sequence of DPPD N-terminal Antigen.

SEQ ID NO. 130-133 are the protein sequences of four DPPD cyanogen bromide fragments.

30 SEQ ID NO. 134 is the N-terminal protein sequence of XDS antigen.

10084843-022502

SEQ ID NO. 135 is the N-terminal protein sequence of AGD antigen.

SEQ ID NO. 136 is the N-terminal protein sequence of APE antigen.

SEQ ID NO. 137 is the N-terminal protein sequence of XYI antigen.

SEQ ID NO. 138 is the DNA sequence of TbH-29.

5 SEQ ID NO. 139 is the DNA sequence of TbH-30.

SEQ ID NO. 140 is the DNA sequence of TbH-32.

SEQ ID NO. 141 is the DNA sequence of TbH-33.

SEQ ID NO. 142 is the predicted amino acid sequence of TbH-29.

SEQ ID NO. 143 is the predicted amino acid sequence of TbH-30.

10 SEQ ID NO. 144 is the predicted amino acid sequence of TbH-32.

SEQ ID NO. 145 is the predicted amino acid sequence of TbH-33.

SEQ ID NO: 146-151 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD and Tb38-1.

15 SEQ ID NO: 152 is the DNA sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 153 is the amino acid sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 154 is the DNA sequence of the *M. tuberculosis* antigen 38 kD.

20 SEQ ID NO: 155 is the amino acid sequence of the *M. tuberculosis* antigen 38 kD.

SEQ ID NO: 156 is the DNA sequence of XP14.

SEQ ID NO: 157 is the DNA sequence of XP24.

SEQ ID NO: 158 is the DNA sequence of XP31.

SEQ ID NO: 159 is the 5' DNA sequence of XP32.

25 SEQ ID NO: 160 is the 3' DNA sequence of XP32.

SEQ ID NO: 161 is the predicted amino acid sequence of XP14.

SEQ ID NO: 162 is the predicted amino acid sequence encoded by the reverse complement of XP14.

SEQ ID NO: 163 is the DNA sequence of XP27.

30 SEQ ID NO: 164 is the DNA sequence of XP36.

205220" E-18-1800T

SEQ ID NO: 165 is the 5' DNA sequence of XP4.

SEQ ID NO: 166 is the 5' DNA sequence of XP5.

SEQ ID NO: 167 is the 5' DNA sequence of XP17.

SEQ ID NO: 168 is the 5' DNA sequence of XP30.

5 SEQ ID NO: 169 is the 5' DNA sequence of XP2.

SEQ ID NO: 170 is the 3' DNA sequence of XP2.

SEQ ID NO: 171 is the 5' DNA sequence of XP3.

SEQ ID NO: 172 is the 3' DNA sequence of XP3.

SEQ ID NO: 173 is the 5' DNA sequence of XP6.

10 SEQ ID NO: 174 is the 3' DNA sequence of XP6.

SEQ ID NO: 175 is the 5' DNA sequence of XP18.

SEQ ID NO: 176 is the 3' DNA sequence of XP18.

SEQ ID NO: 177 is the 5' DNA sequence of XP19.

SEQ ID NO: 178 is the 3' DNA sequence of XP19.

15 SEQ ID NO: 179 is the 5' DNA sequence of XP22.

SEQ ID NO: 180 is the 3' DNA sequence of XP22.

SEQ ID NO: 181 is the 5' DNA sequence of XP25.

SEQ ID NO: 182 is the 3' DNA sequence of XP25.

SEQ ID NO: 183 is the full-length DNA sequence of TbH4-XP1.

20 SEQ ID NO: 184 is the predicted amino acid sequence of TbH4-XP1.

SEQ ID NO: 185 is the predicted amino acid sequence encoded by the reverse complement of TbH4-XP1.

SEQ ID NO: 186 is a first predicted amino acid sequence encoded by XP36.

SEQ ID NO: 187 is a second predicted amino acid sequence encoded by XP36.

25 SEQ ID NO: 188 is the predicted amino acid sequence encoded by the reverse complement of XP36.

SEQ ID NO: 189 is the DNA sequence of RDIF2.

SEQ ID NO: 190 is the DNA sequence of RDIF5.

SEQ ID NO: 191 is the DNA sequence of RDIF8.

30 SEQ ID NO: 192 is the DNA sequence of RDIF10.

10084843.022502

SEQ ID NO: 193 is the DNA sequence of RDIF11.

SEQ ID NO: 194 is the predicted amino acid sequence of RDIF2.

SEQ ID NO: 195 is the predicted amino acid sequence of RDIF5.

SEQ ID NO: 196 is the predicted amino acid sequence of RDIF8.

5 SEQ ID NO: 197 is the predicted amino acid sequence of RDIF10.

SEQ ID NO: 198 is the predicted amino acid sequence of RDIF11.

SEQ ID NO: 199 is the 5' DNA sequence of RDIF12.

SEQ ID NO: 200 is the 3' DNA sequence of RDIF12.

SEQ ID NO: 201 is the DNA sequence of RDIF7.

10 SEQ ID NO: 202 is the predicted amino acid sequence of RDIF7.

SEQ ID NO: 203 is the DNA sequence of DIF2-1.

SEQ ID NO: 204 is the predicted amino acid sequence of DIF2-1.

15 SEQ ID NO: 205-212 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD, Tb38-1 and DPEP (hereinafter referred to as TbF-2).

SEQ ID NO: 213 is the DNA sequence of the fusion protein TbF-2.

SEQ ID NO: 214 is the amino acid sequence of the fusion protein TbF-2.

SEQ ID NO: 215 is the 5' DNA sequence of MO-1.

SEQ ID NO: 216 is the 5' DNA sequence for MO-2

20 SEQ ID NO: 217 is the 5' DNA sequence for MO-4.

SEQ ID NO: 218 is the 5' DNA sequence for MO-8.

SEQ ID NO: 219 is the 5' DNA sequence for MO-9.

SEQ ID NO: 220 is the 5' DNA sequence for MO-26.

SEQ ID NO: 221 is the 5' DNA sequence for MO-28.

25 SEQ ID NO: 222 is the 5' DNA sequence for MO-29.

SEQ ID NO: 223 is the 5' DNA sequence for MO-30.

SEQ ID NO: 224 is the 5' DNA sequence for MO-34.

SEQ ID NO: 225 is the 5' DNA sequence for MO-35.

SEQ ID NO: 226 is the predicted amino acid sequence for MO-1.

30 SEQ ID NO: 227 is the predicted amino acid sequence for MO-2.

10084843-022502

SEQ ID NO: 228 is the predicted amino acid sequence for MO-4.  
SEQ ID NO: 229 is the predicted amino acid sequence for MO-8.  
SEQ ID NO: 230 is the predicted amino acid sequence for MO-9.  
SEQ ID NO: 231 is the predicted amino acid sequence for MO-26.  
5 SEQ ID NO: 232 is the predicted amino acid sequence for MO-28.  
SEQ ID NO: 233 is the predicted amino acid sequence for MO-29.  
SEQ ID NO: 234 is the predicted amino acid sequence for MO-30.  
SEQ ID NO: 235 is the predicted amino acid sequence for MO-34.  
SEQ ID NO: 236 is the predicted amino acid sequence for MO-35.  
10 SEQ ID NO: 237 is the determined DNA sequence for MO-10.  
SEQ ID NO: 238 is the predicted amino acid sequence for MO-10.  
SEQ ID NO: 239 is the 3' DNA sequence for MO-27.  
SEQ ID NO: 240 is the full-length DNA sequence for DPPD.  
SEQ ID NO: 241 is the predicted full-length amino acid sequence for DPPD.  
15 SEQ ID NO: 242 is the determined 5' cDNA sequence for LSER-10  
SEQ ID NO: 243 is the determined 5' cDNA sequence for LSER-11  
SEQ ID NO: 244 is the determined 5' cDNA sequence for LSER-12  
SEQ ID NO: 245 is the determined 5' cDNA sequence for LSER-13  
SEQ ID NO: 246 is the determined 5' cDNA sequence for LSER-16  
20 SEQ ID NO: 247 is the determined 5' cDNA sequence for LSER-25  
SEQ ID NO: 248 is the predicted amino acid sequence for LSER-10  
SEQ ID NO: 249 is the predicted amino acid sequence for LSER-12  
SEQ ID NO: 250 is the predicted amino acid sequence for LSER-13  
SEQ ID NO: 251 is the predicted amino acid sequence for LSER-16  
25 SEQ ID NO: 252 is the predicted amino acid sequence for LSER-25  
SEQ ID NO: 253 is the determined cDNA sequence for LSER-18  
SEQ ID NO: 254 is the determined cDNA sequence for LSER-23  
SEQ ID NO: 255 is the determined cDNA sequence for LSER-24  
SEQ ID NO: 256 is the determined cDNA sequence for LSER-27  
30 SEQ ID NO: 257 is the predicted amino acid sequence for LSER-18

10084843.022502

SEQ ID NO: 258 is the predicted amino acid sequence for LSER-23  
SEQ ID NO: 259 is the predicted amino acid sequence for LSER-24  
SEQ ID NO: 260 is the predicted amino acid sequence for LSER-27  
SEQ ID NO: 261 is the determined 5' cDNA sequence for LSER-1  
5 SEQ ID NO: 262 is the determined 5' cDNA sequence for LSER-3  
SEQ ID NO: 263 is the determined 5' cDNA sequence for LSER-4  
SEQ ID NO: 264 is the determined 5' cDNA sequence for LSER-5  
SEQ ID NO: 265 is the determined 5' cDNA sequence for LSER-6  
SEQ ID NO: 266 is the determined 5' cDNA sequence for LSER-8  
10 SEQ ID NO: 267 is the determined 5' cDNA sequence for LSER-14  
SEQ ID NO: 268 is the determined 5' cDNA sequence for LSER-15  
SEQ ID NO: 269 is the determined 5' cDNA sequence for LSER-17  
SEQ ID NO: 270 is the determined 5' cDNA sequence for LSER-19  
SEQ ID NO: 271 is the determined 5' cDNA sequence for LSER-20  
15 SEQ ID NO: 272 is the determined 5' cDNA sequence for LSER-22  
SEQ ID NO: 273 is the determined 5' cDNA sequence for LSER-26  
SEQ ID NO: 274 is the determined 5' cDNA sequence for LSER-28  
SEQ ID NO: 275 is the determined 5' cDNA sequence for LSER-29  
SEQ ID NO: 276 is the determined 5' cDNA sequence for LSER-30  
20 SEQ ID NO: 277 is the predicted amino acid sequence for LSER-1  
SEQ ID NO: 278 is the predicted amino acid sequence for LSER-3  
SEQ ID NO: 279 is the predicted amino acid sequence for LSER-5  
SEQ ID NO: 280 is the predicted amino acid sequence for LSER-6  
SEQ ID NO: 281 is the predicted amino acid sequence for LSER-8  
25 SEQ ID NO: 282 is the predicted amino acid sequence for LSER-14  
SEQ ID NO: 283 is the predicted amino acid sequence for LSER-15  
SEQ ID NO: 284 is the predicted amino acid sequence for LSER-17  
SEQ ID NO: 285 is the predicted amino acid sequence for LSER-19  
SEQ ID NO: 286 is the predicted amino acid sequence for LSER-20  
30 SEQ ID NO: 287 is the predicted amino acid sequence for LSER-22

10084843.022502



SEQ ID NO: 288 is the predicted amino acid sequence for LSER-26

SEQ ID NO: 289 is the predicted amino acid sequence for LSER-28

SEQ ID NO: 290 is the predicted amino acid sequence for LSER-29

SEQ ID NO: 291 is the predicted amino acid sequence for LSER-30

5 SEQ ID NO: 292 is the determined cDNA sequence for LSER-9

SEQ ID NO: 293 is the determined cDNA sequence for the reverse complement of LSER-6

SEQ ID NO: 294 is the predicted amino acid sequence for the reverse complement of LSER-6

10 SEQ ID NO: 295 is the determined 5' cDNA sequence for MO-12

SEQ ID NO: 296 is the determined 5' cDNA sequence for MO-13

SEQ ID NO: 297 is the determined 5' cDNA sequence for MO-19

SEQ ID NO: 298 is the determined 5' cDNA sequence for MO-39

SEQ ID NO: 299 is the predicted amino acid sequence for MO-12

15 SEQ ID NO: 300 is the predicted amino acid sequence for MO-13

SEQ ID NO: 301 is the predicted amino acid sequence for MO-19

SEQ ID NO: 302 is the predicted amino acid sequence for MO-39

SEQ ID NO: 303 is the determined 5' cDNA sequence for Erdsn-1

SEQ ID NO: 304 is the determined 5' cDNA sequence for Erdsn-2

20 SEQ ID NO: 305 is the determined 5' cDNA sequence for Erdsn-4

SEQ ID NO: 306 is the determined 5' cDNA sequence for Erdsn-5

SEQ ID NO: 307 is the determined 5' cDNA sequence for Erdsn-6

SEQ ID NO: 308 is the determined 5' cDNA sequence for Erdsn-7

SEQ ID NO: 309 is the determined 5' cDNA sequence for Erdsn-8

25 SEQ ID NO: 310 is the determined 5' cDNA sequence for Erdsn-9

SEQ ID NO: 311 is the determined 5' cDNA sequence for Erdsn-10

SEQ ID NO: 312 is the determined 5' cDNA sequence for Erdsn-12

SEQ ID NO: 313 is the determined 5' cDNA sequence for Erdsn-13

SEQ ID NO: 314 is the determined 5' cDNA sequence for Erdsn-14

30 SEQ ID NO: 315 is the determined 5' cDNA sequence for Erdsn-15

10084843 02502

SEQ ID NO: 316 is the determined 5' cDNA sequence for Erdsn-16  
 SEQ ID NO: 317 is the determined 5' cDNA sequence for Erdsn-17  
 SEQ ID NO: 318 is the determined 5' cDNA sequence for Erdsn-18  
 SEQ ID NO: 319 is the determined 5' cDNA sequence for Erdsn-21  
 5 SEQ ID NO: 320 is the determined 5' cDNA sequence for Erdsn-22  
 SEQ ID NO: 321 is the determined 5' cDNA sequence for Erdsn-23  
 SEQ ID NO: 322 is the determined 5' cDNA sequence for Erdsn-25  
 SEQ ID NO: 323 is the determined 3' cDNA sequence for Erdsn-1  
 SEQ ID NO: 324 is the determined 3' cDNA sequence for Erdsn-2  
 10 SEQ ID NO: 325 is the determined 3' cDNA sequence for Erdsn-4  
 SEQ ID NO: 326 is the determined 3' cDNA sequence for Erdsn-5  
 SEQ ID NO: 327 is the determined 3' cDNA sequence for Erdsn-7  
 SEQ ID NO: 328 is the determined 3' cDNA sequence for Erdsn-8  
 SEQ ID NO: 329 is the determined 3' cDNA sequence for Erdsn-9  
 15 SEQ ID NO: 330 is the determined 3' cDNA sequence for Erdsn-10  
 SEQ ID NO: 331 is the determined 3' cDNA sequence for Erdsn-12  
 SEQ ID NO: 332 is the determined 3' cDNA sequence for Erdsn-13  
 SEQ ID NO: 333 is the determined 3' cDNA sequence for Erdsn-14  
 SEQ ID NO: 334 is the determined 3' cDNA sequence for Erdsn-15  
 20 SEQ ID NO: 335 is the determined 3' cDNA sequence for Erdsn-16  
 SEQ ID NO: 336 is the determined 3' cDNA sequence for Erdsn-17  
 SEQ ID NO: 337 is the determined 3' cDNA sequence for Erdsn-18  
 SEQ ID NO: 338 is the determined 3' cDNA sequence for Erdsn-21  
 SEQ ID NO: 339 is the determined 3' cDNA sequence for Erdsn-22  
 25 SEQ ID NO: 340 is the determined 3' cDNA sequence for Erdsn-23  
 SEQ ID NO: 341 is the determined 3' cDNA sequence for Erdsn-25  
 SEQ ID NO: 342 is the determined cDNA sequence for Erdsn-24  
 SEQ ID NO: 343 is the determined amino acid sequence for a *M. tuberculosis*  
 85b precursor homolog  
 30 SEQ ID NO: 344 is the determined amino acid sequence for spot 1

205220 "E484800T

SEQ ID NO: 345 is a determined amino acid sequence for spot 2

SEQ ID NO: 346 is a determined amino acid sequence for spot 2

SEQ ID NO: 347 is the determined amino acid seq for spot 4

SEQ ID NO: 348 is the sequence of primer PDM-157

5 SEQ ID NO: 349 is the sequence of primer PDM-160

SEQ ID NO: 350 is the DNA sequence of the fusion protein TbF-6

SEQ ID NO: 351 is the amino acid sequence of fusion protein TbF-6

SEQ ID NO: 352 is the sequence of primer PDM-176

SEQ ID NO: 353 is the sequence of primer PDM-175

10 SEQ ID NO: 354 is the DNA sequence of the fusion protein TbF-8

SEQ ID NO: 355 is the amino acid sequence of the fusion protein TbF-8

#### DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is generally directed to  
 15 compositions and methods for preventing, treating and diagnosing tuberculosis. The  
 compositions of the subject invention include polypeptides that comprise at least one  
 immunogenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that  
 differs only in conservative substitutions and/or modifications. Polypeptides within the  
 scope of the present invention include, but are not limited to, immunogenic soluble  
 20 *M. tuberculosis* antigens. A "soluble *M. tuberculosis* antigen" is a protein of  
*M. tuberculosis* origin that is present in *M. tuberculosis* culture filtrate. As used herein,  
 the term "polypeptide" encompasses amino acid chains of any length, including full  
 length proteins (*i.e.*, antigens), wherein the amino acid residues are linked by covalent  
 peptide bonds. Thus, a polypeptide comprising an immunogenic portion of one of the  
 25 above antigens may consist entirely of the immunogenic portion, or may contain  
 additional sequences. The additional sequences may be derived from the native  
*M. tuberculosis* antigen or may be heterologous, and such sequences may (but need not)  
 be immunogenic.

"Immunogenic," as used herein, refers to the ability to elicit an immune  
 30 response (*e.g.*, cellular) in a patient, such as a human, and/or in a biological sample. In  
 particular, antigens that are immunogenic (and immunogenic portions or other variants

10084843.022502  
205220.488001

of such antigens) are capable of stimulating cell proliferation, interleukin-12 production and/or interferon- $\gamma$  production in biological samples comprising one or more cells selected from the group of T cells, NK cells, B cells and macrophages, where the cells are derived from an *M. tuberculosis*-immune individual. Polypeptides comprising at least an immunogenic portion of one or more *M. tuberculosis* antigens may generally be used to detect tuberculosis or to induce protective immunity against tuberculosis in a patient.

The compositions and methods of the present invention also encompass variants of the above polypeptides and DNA molecules. A polypeptide "variant," as used herein, is a polypeptide that differs from the recited polypeptide only in conservative substitutions and/or modifications, such that the therapeutic, antigenic and/or immunogenic properties of the polypeptide are retained. Polypeptide variants preferably exhibit at least about 70%, more preferably at least about 90% and most preferably at least about 95% identity to the identified polypeptides. For polypeptides with immunoreactive properties, variants may, alternatively, be identified by modifying the amino acid sequence of one of the above polypeptides, and evaluating the immunoreactivity of the modified polypeptide. For polypeptides useful for the generation of diagnostic binding agents, a variant may be identified by evaluating a modified polypeptide for the ability to generate antibodies that detect the presence or absence of tuberculosis. Such modified sequences may be prepared and tested using, for example, the representative procedures described herein.

As used herein, a "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his.

Variants may also, or alternatively, contain other modifications, including the deletion or addition of amino acids that have minimal influence on the

10084843.022502

antigenic properties, secondary structure and hydropathic nature of the polypeptide. For example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other  
 5 sequence for ease of synthesis, purification or identification of the polypeptide (e.g., poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

A nucleotide "variant" is a sequence that differs from the recited nucleotide sequence in having one or more nucleotide deletions, substitutions or  
 10 additions. Such modifications may be readily introduced using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis as taught, for example, by Adelman et al. (*DNA*, 2:183, 1983). Nucleotide variants may be naturally occurring allelic variants, or non-naturally occurring variants. Variant nucleotide sequences preferably exhibit at least about 70%, more preferably at least about 80% and  
 15 most preferably at least about 90% identity to the recited sequence. Such variant nucleotide sequences will generally hybridize to the recite nucleotide sequence under stringent conditions. As used herein, "stringent conditions" refers to prewashing in a solution of 6X SSC, 0.2% SDS; hybridizing at 65 °C, 6X SSC, 0.2% SDS overnight; followed by two washes of 30 minutes each in 1X SSC, 0.1% SDS at 65 °C and two  
 20 washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65 °C.

In a related aspect, combination polypeptides are disclosed. A "combination polypeptide" is a polypeptide comprising at least one of the above immunogenic portions and one or more additional immunogenic *M. tuberculosis* sequences, which are joined via a peptide linkage into a single amino acid chain. The  
 25 sequences may be joined directly (i.e., with no intervening amino acids) or may be joined by way of a linker sequence (e.g., Gly-Cys-Gly) that does not significantly diminish the immunogenic properties of the component polypeptides.

In general, *M. tuberculosis* antigens, and DNA sequences encoding such antigens, may be prepared using any of a variety of procedures. For example, soluble  
 30 antigens may be isolated from *M. tuberculosis* culture filtrate by procedures known to

those of ordinary skill in the art, including anion-exchange and reverse phase chromatography. Purified antigens are then evaluated for their ability to elicit an appropriate immune response (e.g., cellular) using, for example, the representative methods described herein. Immunogenic antigens may then be partially sequenced  
5 using techniques such as traditional Edman chemistry. See Edman and Berg, *Eur. J. Biochem.* 80:116-132, 1967.

Immunogenic antigens may also be produced recombinantly using a DNA sequence that encodes the antigen, which has been inserted into an expression vector and expressed in an appropriate host. DNA molecules encoding soluble antigens  
10 may be isolated by screening an appropriate *M. tuberculosis* expression library with anti-sera (e.g., rabbit) raised specifically against soluble *M. tuberculosis* antigens. DNA sequences encoding antigens that may or may not be soluble may be identified by screening an appropriate *M. tuberculosis* genomic or cDNA expression library with sera  
15 obtained from patients infected with *M. tuberculosis*. Such screens may generally be performed using techniques well known to those of ordinary skill in the art, such as those described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989.

DNA sequences encoding soluble antigens may also be obtained by screening an appropriate *M. tuberculosis* cDNA or genomic DNA library for DNA  
20 sequences that hybridize to degenerate oligonucleotides derived from partial amino acid sequences of isolated soluble antigens. Degenerate oligonucleotide sequences for use in such a screen may be designed and synthesized, and the screen may be performed, as described (for example) in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989 (and references cited  
25 therein). Polymerase chain reaction (PCR) may also be employed, using the above oligonucleotides in methods well known in the art, to isolate a nucleic acid probe from a cDNA or genomic library. The library screen may then be performed using the isolated probe.

Alternatively, genomic or cDNA libraries derived from *M. tuberculosis*  
30 may be screened directly using peripheral blood mononuclear cells (PBMCs) or T cell

205220" CH84800T

lines or clones derived from one or more *M. tuberculosis*-immune individuals. In general, PBMCs and/or T cells for use in such screens may be prepared as described below. Direct library screens may generally be performed by assaying pools of expressed recombinant proteins for the ability to induce proliferation and/or interferon- $\gamma$  production in T cells derived from an *M. tuberculosis*-immune individual. Alternatively, potential T cell antigens may be first selected based on antibody reactivity, as described above.

Regardless of the method of preparation, the antigens (and immunogenic portions thereof) described herein (which may or may not be soluble) have the ability to induce an immunogenic response. More specifically, the antigens have the ability to induce proliferation and/or cytokine production (*i.e.*, interferon- $\gamma$  and/or interleukin-12 production) in T cells, NK cells, B cells and/or macrophages derived from an *M. tuberculosis*-immune individual. The selection of cell type for use in evaluating an immunogenic response to a antigen will, of course, depend on the desired response. For example, interleukin-12 production is most readily evaluated using preparations containing B cells and/or macrophages. An *M. tuberculosis*-immune individual is one who is considered to be resistant to the development of tuberculosis by virtue of having mounted an effective T cell response to *M. tuberculosis* (*i.e.*, substantially free of disease symptoms). Such individuals may be identified based on a strongly positive (*i.e.*, greater than about 10 mm diameter induration) intradermal skin test response to tuberculosis proteins (PPD) and an absence of any signs or symptoms of tuberculosis disease. T cells, NK cells, B cells and macrophages derived from *M. tuberculosis*-immune individuals may be prepared using methods known to those of ordinary skill in the art. For example, a preparation of PBMCs (*i.e.*, peripheral blood mononuclear cells) may be employed without further separation of component cells. PBMCs may generally be prepared, for example, using density centrifugation through Ficoll™ (Winthrop Laboratories, NY). T cells for use in the assays described herein may also be purified directly from PBMCs. Alternatively, an enriched T cell line reactive against mycobacterial proteins, or T cell clones reactive to individual mycobacterial proteins, may be employed. Such T cell clones may be generated by, for example, culturing

PBMCs from *M. tuberculosis*-immune individuals with mycobacterial proteins for a period of 2-4 weeks. This allows expansion of only the mycobacterial protein-specific T cells, resulting in a line composed solely of such cells. These cells may then be cloned and tested with individual proteins, using methods known to those of ordinary skill in the art, to more accurately define individual T cell specificity. In general, antigens that test positive in assays for proliferation and/or cytokine production (*i.e.*, interferon- $\gamma$  and/or interleukin-12 production) performed using T cells, NK cells, B cells and/or macrophages derived from an *M. tuberculosis*-immune individual are considered immunogenic. Such assays may be performed, for example, using the representative procedures described below. Immunogenic portions of such antigens may be identified using similar assays, and may be present within the polypeptides described herein.

The ability of a polypeptide (*e.g.*, an immunogenic antigen, or a portion or other variant thereof) to induce cell proliferation is evaluated by contacting the cells (*e.g.*, T cells and/or NK cells) with the polypeptide and measuring the proliferation of the cells. In general, the amount of polypeptide that is sufficient for evaluation of about  $10^5$  cells ranges from about 10 ng/mL to about 100  $\mu$ g/mL and preferably is about 10  $\mu$ g/mL. The incubation of polypeptide with cells is typically performed at 37°C for about six days. Following incubation with polypeptide, the cells are assayed for a proliferative response, which may be evaluated by methods known to those of ordinary skill in the art, such as exposing cells to a pulse of radiolabeled thymidine and measuring the incorporation of label into cellular DNA. In general, a polypeptide that results in at least a three fold increase in proliferation above background (*i.e.*, the proliferation observed for cells cultured without polypeptide) is considered to be able to induce proliferation.

The ability of a polypeptide to stimulate the production of interferon- $\gamma$  and/or interleukin-12 in cells may be evaluated by contacting the cells with the polypeptide and measuring the level of interferon- $\gamma$  or interleukin-12 produced by the cells. In general, the amount of polypeptide that is sufficient for the evaluation of about  $10^5$  cells ranges from about 10 ng/mL to about 100  $\mu$ g/mL and preferably is about 10  $\mu$ g/mL. The polypeptide may, but need not, be immobilized on a solid support, such as a

205220" CH848001



bead or a biodegradable microsphere, such as those described in U.S. Patent Nos. 4,897,268 and 5,075,109. The incubation of polypeptide with the cells is typically performed at 37°C for about six days. Following incubation with polypeptide, the cells are assayed for interferon- $\gamma$  and/or interleukin-12 (or one or more subunits thereof), which may be evaluated by methods known to those of ordinary skill in the art, such as an enzyme-linked immunosorbent assay (ELISA) or, in the case of IL-12 P70 subunit, a bioassay such as an assay measuring proliferation of T cells. In general, a polypeptide that results in the production of at least 50 pg of interferon- $\gamma$  per mL of cultured supernatant (containing  $10^4$ - $10^5$  T cells per mL) is considered able to stimulate the production of interferon- $\gamma$ . A polypeptide that stimulates the production of at least 10 pg/mL of IL-12 P70 subunit, and/or at least 100 pg/mL of IL-12 P40 subunit, per  $10^5$  macrophages or B cells (or per  $3 \times 10^5$  PBMC) is considered able to stimulate the production of IL-12.

In general, immunogenic antigens are those antigens that stimulate proliferation and/or cytokine production (i.e., interferon- $\gamma$  and/or interleukin-12 production) in T cells, NK cells, B cells and/or macrophages derived from at least about 25% of *M. tuberculosis*-immune individuals. Among these immunogenic antigens, polypeptides having superior therapeutic properties may be distinguished based on the magnitude of the responses in the above assays and based on the percentage of individuals for which a response is observed. In addition, antigens having superior therapeutic properties will not stimulate proliferation and/or cytokine production *in vitro* in cells derived from more than about 25% of individuals that are not *M. tuberculosis*-immune, thereby eliminating responses that are not specifically due to *M. tuberculosis*-responsive cells. Those antigens that induce a response in a high percentage of T cell, NK cell, B cell and/or macrophage preparations from *M. tuberculosis*-immune individuals (with a low incidence of responses in cell preparations from other individuals) have superior therapeutic properties.

Antigens with superior therapeutic properties may also be identified based on their ability to diminish the severity of *M. tuberculosis* infection in experimental animals, when administered as a vaccine. Suitable vaccine preparations

for use on experimental animals are described in detail below. Efficacy may be determined based on the ability of the antigen to provide at least about a 50% reduction in bacterial numbers and/or at least about a 40% decrease in mortality following experimental infection. Suitable experimental animals include mice, guinea pigs and  
5 primates.

Antigens having superior diagnostic properties may generally be identified based on the ability to elicit a response in an intradermal skin test performed on an individual with active tuberculosis, but not in a test performed on an individual who is not infected with *M. tuberculosis*. Skin tests may generally be performed as  
10 described below, with a response of at least 5 mm induration considered positive.

Immunogenic portions of the antigens described herein may be prepared and identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3d ed., Raven Press, 1993, pp. 243-247 and references cited therein. Such techniques include screening polypeptide portions of the native antigen  
15 for immunogenic properties. The representative proliferation and cytokine production assays described herein may generally be employed in these screens. An immunogenic portion of a polypeptide is a portion that, within such representative assays, generates an immune response (e.g., proliferation, interferon- $\gamma$  production and/or interleukin-12 production) that is substantially similar to that generated by the full length antigen. In  
20 other words, an immunogenic portion of an antigen may generate at least about 20%, and preferably about 100%, of the proliferation induced by the full length antigen in the model proliferation assay described herein. An immunogenic portion may also, or alternatively, stimulate the production of at least about 20%, and preferably about 100%, of the interferon- $\gamma$  and/or interleukin-12 induced by the full length antigen in the  
25 model assay described herein.

Portions and other variants of *M. tuberculosis* antigens may be generated by synthetic or recombinant means. Synthetic polypeptides having fewer than about 100 amino acids, and generally fewer than about 50 amino acids, may be generated using techniques well known to those of ordinary skill in the art. For example, such  
30 polypeptides may be synthesized using any of the commercially available solid-phase

10084843 1022502 2052201 CH98007

techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, *J. Am. Chem. Soc.* 85:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Applied BioSystems, Inc., Foster City, CA, and may be operated according to the manufacturer's instructions. Variants of a native antigen may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis. Sections of the DNA sequence may also be removed using standard techniques to permit preparation of truncated polypeptides.

Recombinant polypeptides containing portions and/or variants of a native antigen may be readily prepared from a DNA sequence encoding the polypeptide using a variety of techniques well known to those of ordinary skill in the art. For example, supernatants from suitable host/vector systems which secrete recombinant protein into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant protein.

Any of a variety of expression vectors known to those of ordinary skill in the art may be employed to express recombinant polypeptides of this invention. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO. The DNA sequences expressed in this manner may encode naturally occurring antigens, portions of naturally occurring antigens, or other variants thereof.

In general, regardless of the method of preparation, the polypeptides disclosed herein are prepared in substantially pure form. Preferably, the polypeptides are at least about 80% pure, more preferably at least about 90% pure and most

10084843.022502

preferably at least about 99% pure. In certain preferred embodiments, described in detail below, the substantially pure polypeptides are incorporated into pharmaceutical compositions or vaccines for use in one or more of the methods disclosed herein.

In certain specific embodiments, the subject invention discloses  
 5 polypeptides comprising at least an immunogenic portion of a soluble *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only in conservative substitutions and/or modifications:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 120)
- 10 (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 123)
- 15 (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 124)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 125)
- 20 (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ser-Pro-Pro-Ser; (SEQ ID No. 126)
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn; (SEQ ID No. 128)
- 25 (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 134)
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) or
- 30

205220" CH84800T  
 10084843 022502

- (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

wherein Xaa may be any amino acid, preferably a cysteine residue. A DNA sequence encoding the antigen identified as (g) above is provided in SEQ ID No. 52, and the polypeptide encoded by SEQ ID No. 52 is provided in SEQ ID No. 53. A DNA sequence encoding the antigen defined as (a) above is provided in SEQ ID No. 101; its deduced amino acid sequence is provided in SEQ ID No. 102. A DNA sequence corresponding to antigen (d) above is provided in SEQ ID No. 24 a DNA sequence corresponding to antigen (c) is provided in SEQ ID No. 25 and a DNA sequence corresponding to antigen (i) is provided in SEQ ID No. 99; its deduced amino acid sequence is provided in SEQ ID No. 100.

In a further specific embodiment, the subject invention discloses polypeptides comprising at least an immunogenic portion of an *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only in conservative substitutions and/or modifications:

- (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No 137) or  
(n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID No. 129)

wherein Xaa may be any amino acid, preferably a cysteine residue.

In other specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a soluble *M. tuberculosis* antigen (or a variant of such an antigen) that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID Nos.: 1, 2, 4-10, 13-25 and 52; (b) the complements of such DNA sequences, or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

In further specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a *M. tuberculosis* antigen (or a variant of such an antigen), which may or may not be soluble, that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID

10084843 022502

Nos.: 26-51, 138, 139, 163-183, 189-193, 199, 200, 201, 203, 215-225, 239, 240, 242-247, 253-256, 261-276, 292, 293, 295-298 and 303-342, (b) the complements of such DNA sequences or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

5 In the specific embodiments discussed above, the *M. tuberculosis* antigens include variants that are encoded by DNA sequences which are substantially homologous to one or more of DNA sequences specifically recited herein. "Substantial homology," as used herein, refers to DNA sequences that are capable of hybridizing under moderately stringent conditions. Suitable moderately stringent conditions include  
10 prewashing in a solution of 5X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5X SSC, overnight or, in the case of cross-species homology at 45°C, 0.5X SSC; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS). Such hybridizing DNA sequences are also within the scope of this invention, as are nucleotide sequences that, due to code  
15 degeneracy, encode an immunogenic polypeptide that is encoded by a hybridizing DNA sequence.

In a related aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, a polypeptide of the present invention and a known *M. tuberculosis* antigen, such as the 38 kD antigen  
20 described in Andersen and Hansen, *Infect. Immun.* 57:2481-2488, 1989, (Genbank Accession No. M30046) or ESAT-6 (SEQ ID Nos. 103 and 104), together with variants of such fusion proteins. The fusion proteins of the present invention may also include a linker peptide between the first and second polypeptides.

A DNA sequence encoding a fusion protein of the present invention is  
25 constructed using known recombinant DNA techniques to assemble separate DNA sequences encoding the first and second polypeptides into an appropriate expression vector. The 3' end of a DNA sequence encoding the first polypeptide is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide so that the reading frames of the sequences are in phase to permit mRNA

10084843 022502

translation of the two DNA sequences into a single fusion protein that retains the biological activity of both the first and the second polypeptides.

A peptide linker sequence may be employed to separate the first and the second polypeptides by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene* 40:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may be from 1 to about 50 amino acids in length. Peptide sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons require to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

In another aspect, the present invention provides methods for using one or more of the above polypeptides or fusion proteins (or DNA molecules encoding such polypeptides) to induce protective immunity against tuberculosis in a patient. As used herein, a "patient" refers to any warm-blooded animal, preferably a human. A patient

10084843 "022502

may be afflicted with a disease, or may be free of detectable disease and/or infection. In other words, protective immunity may be induced to prevent or treat tuberculosis.

In this aspect, the polypeptide, fusion protein or DNA molecule is generally present within a pharmaceutical composition and/or a vaccine.

5 Pharmaceutical compositions may comprise one or more polypeptides, each of which may contain one or more of the above sequences (or variants thereof), and a physiologically acceptable carrier. Vaccines may comprise one or more of the above polypeptides and a non-specific immune response enhancer, such as an adjuvant or a liposome (into which the polypeptide is incorporated). Such pharmaceutical

10 compositions and vaccines may also contain other *M. tuberculosis* antigens, either incorporated into a combination polypeptide or present within a separate polypeptide.

Alternatively, a vaccine may contain DNA encoding one or more polypeptides as described above, such that the polypeptide is generated *in situ*. In such vaccines, the DNA may be present within any of a variety of delivery systems known to

15 those of ordinary skill in the art, including nucleic acid expression systems, bacterial and viral expression systems. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion

20 of the polypeptide on its cell surface. In a preferred embodiment, the DNA may be introduced using a viral expression system (e.g., vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication competent virus. Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art. The DNA may also be "naked," as

25 described, for example, in Ulmer et al., *Science* 259:1745-1749, 1993 and reviewed by Cohen, *Science* 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

In a related aspect, a DNA vaccine as described above may be

30 administered simultaneously with or sequentially to either a polypeptide of the present

10084843-022502



invention or a known *M. tuberculosis* antigen, such as the 38 kD antigen described above. For example, administration of DNA encoding a polypeptide of the present invention, either "naked" or in a delivery system as described above, may be followed by administration of an antigen in order to enhance the protective immune effect of the vaccine.

Routes and frequency of administration, as well as dosage, will vary from individual to individual and may parallel those currently being used in immunization using BCG. In general, the pharmaceutical compositions and vaccines may be administered by injection (*e.g.*, intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (*e.g.*, by aspiration) or orally. Between 1 and 3 doses may be administered for a 1-36 week period. Preferably, 3 doses are administered, at intervals of 3-4 months, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of polypeptide or DNA that, when administered as described above, is capable of raising an immune response in an immunized patient sufficient to protect the patient from *M. tuberculosis* infection for at least 1-2 years. In general, the amount of polypeptide present in a dose (or produced *in situ* by the DNA in a dose) ranges from about 1 pg to about 100 mg per kg of host, typically from about 10 pg to about 1 mg, and preferably from about 100 pg to about 1 µg. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will vary depending on the mode of administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier, such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (*e.g.*, polylactic galactide) may also be employed as carriers for the pharmaceutical compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268 and 5,075,109.

Any of a variety of adjuvants may be employed in the vaccines of this invention to nonspecifically enhance the immune response. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a nonspecific stimulator of immune responses, such as lipid A, *Bordetella pertussis* or *Mycobacterium tuberculosis*. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Freund's Complete Adjuvant (Difco Laboratories) and Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ). Other suitable adjuvants include alum, biodegradable microspheres, monophosphoryl lipid A and quil A.

In another aspect, this invention provides methods for using one or more of the polypeptides described above to diagnose tuberculosis using a skin test. As used herein, a "skin test" is any assay performed directly on a patient in which a delayed-type hypersensitivity (DTH) reaction (such as swelling, reddening or dermatitis) is measured following intradermal injection of one or more polypeptides as described above. Such injection may be achieved using any suitable device sufficient to contact the polypeptide or polypeptides with dermal cells of the patient, such as a tuberculin syringe or 1 mL syringe. Preferably, the reaction is measured at least 48 hours after injection, more preferably 48-72 hours.

The DTH reaction is a cell-mediated immune response, which is greater in patients that have been exposed previously to the test antigen (*i.e.*, the immunogenic portion of the polypeptide employed, or a variant thereof). The response may be measured visually, using a ruler. In general, a response that is greater than about 0.5 cm in diameter, preferably greater than about 1.0 cm in diameter, is a positive response, indicative of tuberculosis infection, which may or may not be manifested as an active disease.

The polypeptides of this invention are preferably formulated, for use in a skin test, as pharmaceutical compositions containing a polypeptide and a physiologically acceptable carrier, as described above. Such compositions typically contain one or more of the above polypeptides in an amount ranging from about 1  $\mu$ g to about 100  $\mu$ g, preferably from about 10  $\mu$ g to about 50  $\mu$ g in a volume of 0.1 mL.

10084843 "022502

Preferably, the carrier employed in such pharmaceutical compositions is a saline solution with appropriate preservatives, such as phenol and/or Tween 80™.

In a preferred embodiment, a polypeptide employed in a skin test is of sufficient size such that it remains at the site of injection for the duration of the reaction period. In general, a polypeptide that is at least 9 amino acids in length is sufficient. The polypeptide is also preferably broken down by macrophages within hours of injection to allow presentation to T-cells. Such polypeptides may contain repeats of one or more of the above sequences and/or other immunogenic or nonimmunogenic sequences.

The following Examples are offered by way of illustration and not by way of limitation.

## EXAMPLES

### EXAMPLE 1

#### PURIFICATION AND CHARACTERIZATION OF POLYPEPTIDES FROM *M. TUBERCULOSIS* CULTURE FILTRATE

This example illustrates the preparation of *M. tuberculosis* soluble polypeptides from culture filtrate. Unless otherwise noted, all percentages in the following example are weight per volume.

*M. tuberculosis* (either H37Ra, ATCC No. 25177, or H37Rv, ATCC No. 25618) was cultured in sterile GAS media at 37°C for fourteen days. The media was then vacuum filtered (leaving the bulk of the cells) through a 0.45 μ filter into a sterile 2.5 L bottle. The media was next filtered through a 0.2 μ filter into a sterile 4 L bottle and NaN<sub>3</sub> was added to the culture filtrate to a concentration of 0.04%. The bottles were then placed in a 4°C cold room.

The culture filtrate was concentrated by placing the filtrate in a 12 L reservoir that had been autoclaved and feeding the filtrate into a 400 ml Amicon stir cell

which had been rinsed with ethanol and contained a 10,000 kDa MWCO membrane. The pressure was maintained at 60 psi using nitrogen gas. This procedure reduced the 12 L volume to approximately 50 ml.

The culture filtrate was dialyzed into 0.1% ammonium bicarbonate using  
5 a 8,000 kDa MWCO cellulose ester membrane, with two changes of ammonium bicarbonate solution. Protein concentration was then determined by a commercially available BCA assay (Pierce, Rockford, IL).

The dialyzed culture filtrate was then lyophilized, and the polypeptides resuspended in distilled water. The polypeptides were dialyzed against 0.01 mM 1,3  
10 bis[tris(hydroxymethyl)-methylamino]propane, pH 7.5 (Bis-Tris propane buffer), the initial conditions for anion exchange chromatography. Fractionation was performed using gel perfusion chromatography on a POROS 146 II Q/M anion exchange column 4.6 mm x 100 mm (Perseptive BioSystems, Framingham, MA) equilibrated in 0.01 mM Bis-Tris propane buffer pH 7.5. Polypeptides were eluted with a linear 0-0.5 M NaCl  
15 gradient in the above buffer system. The column eluent was monitored at a wavelength of 220 nm.

The pools of polypeptides eluting from the ion exchange column were dialyzed against distilled water and lyophilized. The resulting material was dissolved in 0.1% trifluoroacetic acid (TFA) pH 1.9 in water, and the polypeptides were purified on  
20 a Delta-Pak C18 column (Waters, Milford, MA) 300 Angstrom pore size, 5 micron particle size (3.9 x 150 mm). The polypeptides were eluted from the column with a linear gradient from 0-60% dilution buffer (0.1% TFA in acetonitrile). The flow rate was 0.75 ml/minute and the HPLC eluent was monitored at 214 nm. Fractions containing the eluted polypeptides were collected to maximize the purity of the  
25 individual samples. Approximately 200 purified polypeptides were obtained.

The purified polypeptides were then screened for the ability to induce T-cell proliferation in PBMC preparations. The PBMCs from donors known to be PPD skin test positive and whose T-cells were shown to proliferate in response to PPD and crude soluble proteins from MTB were cultured in medium comprising RPMI 1640  
30 supplemented with 10% pooled human serum and 50 µg/ml gentamicin. Purified

polypeptides were added in duplicate at concentrations of 0.5 to 10  $\mu\text{g/mL}$ . After six days of culture in 96-well round-bottom plates in a volume of 200  $\mu\text{l}$ , 50  $\mu\text{l}$  of medium was removed from each well for determination of IFN- $\gamma$  levels, as described below. The plates were then pulsed with 1  $\mu\text{Ci/well}$  of tritiated thymidine for a further 18  
 5 hours, harvested and tritium uptake determined using a gas scintillation counter. Fractions that resulted in proliferation in both replicates three fold greater than the proliferation observed in cells cultured in medium alone were considered positive.

IFN- $\gamma$  was measured using an enzyme-linked immunosorbent assay (ELISA). ELISA plates were coated with a mouse monoclonal antibody directed to  
 10 human IFN- $\gamma$  (PharMingen, San Diego, CA) in PBS for four hours at room temperature. Wells were then blocked with PBS containing 5% (W/V) non-fat dried milk for 1 hour at room temperature. The plates were then washed six times in PBS/0.2% TWEEN-20 and samples diluted 1:2 in culture medium in the ELISA plates were incubated overnight at room temperature. The plates were again washed and a polyclonal rabbit  
 15 anti-human IFN- $\gamma$  serum diluted 1:3000 in PBS/10% normal goat serum was added to each well. The plates were then incubated for two hours at room temperature, washed and horseradish peroxidase-coupled anti-rabbit IgG (Sigma Chemical So., St. Louis, MO) was added at a 1:2000 dilution in PBS/5% non-fat dried milk. After a further two hour incubation at room temperature, the plates were washed and TMB substrate added.  
 20 The reaction was stopped after 20 min with 1 N sulfuric acid. Optical density was determined at 450 nm using 570 nm as a reference wavelength. Fractions that resulted in both replicates giving an OD two fold greater than the mean OD from cells cultured in medium alone, plus 3 standard deviations, were considered positive.

For sequencing, the polypeptides were individually dried onto Biobrene  
 25 <sup>TM</sup> (Perkin Elmer/Applied BioSystems Division, Foster City, CA) treated glass fiber filters. The filters with polypeptide were loaded onto a Perkin Elmer/Applied BioSystems Division Procise 492 protein sequencer. The polypeptides were sequenced from the amino terminal and using traditional Edman chemistry. The amino acid sequence was determined for each polypeptide by comparing the retention time of the  
 30 PTH amino acid derivative to the appropriate PTH derivative standards.

10084843.022503

Using the procedure described above, antigens having the following N-terminal sequences were isolated:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Xaa-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 54)
- 5 (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 55)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 56)
- 10 (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 57)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 58)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 59)
- 15 (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ala-Pro-Pro-Ala; (SEQ ID No. 60) and
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 61)

wherein Xaa may be any amino acid.

- 20 An additional antigen was isolated employing a microbore HPLC purification step in addition to the procedure described above. Specifically, 20  $\mu$ l of a fraction comprising a mixture of antigens from the chromatographic purification step previously described, was purified on an Aquapore C18 column (Perkin Elmer/Applied Biosystems Division, Foster City, CA) with a 7 micron pore size, column size 1 mm.x
- 25 100 mm, in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted from the column with a linear gradient of 1%/minute of acetonitrile (containing 0.05% TFA) in water (0.05% TFA) at a flow rate of 80  $\mu$ l/minute. The eluent was monitored at 250 nm. The original fraction was separated into 4 major peaks plus other smaller components and a polypeptide was obtained which was shown to

have a molecular weight of 12.054 Kd (by mass spectrometry) and the following N-terminal sequence:

- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-Thr-Ser-Leu-Leu-Asn-Asn-Leu-Ala-Asp-Pro-Asp-Val-Ser-Phe-Ala-Asp (SEQ ID No. 62).

This polypeptide was shown to induce proliferation and IFN- $\gamma$  production in PBMC preparations using the assays described above.

Additional soluble antigens were isolated from *M. tuberculosis* culture filtrate as follows. *M. tuberculosis* culture filtrate was prepared as described above.

Following dialysis against Bis-Tris propane buffer, at pH 5.5, fractionation was performed using anion exchange chromatography on a Poros QE column 4.6 x 100 mm (Perseptive Biosystems) equilibrated in Bis-Tris propane buffer pH 5.5. Polypeptides were eluted with a linear 0-1.5 M NaCl gradient in the above buffer system at a flow rate of 10 ml/min. The column eluent was monitored at a wavelength of 214 nm.

The fractions eluting from the ion exchange column were pooled and subjected to reverse phase chromatography using a Poros R2 column 4.6 x 100 mm (Perseptive Biosystems). Polypeptides were eluted from the column with a linear gradient from 0-100% acetonitrile (0.1% TFA) at a flow rate of 5 ml/min. The eluent was monitored at 214 nm.

Fractions containing the eluted polypeptides were lyophilized and resuspended in 80  $\mu$ l of aqueous 0.1% TFA and further subjected to reverse phase chromatography on a Vydac C4 column 4.6 x 150 mm (Western Analytical, Temecula, CA) with a linear gradient of 0-100% acetonitrile (0.1% TFA) at a flow rate of 2 ml/min. Eluent was monitored at 214 nm.

The fraction with biological activity was separated into one major peak plus other smaller components. Western blot of this peak onto PVDF membrane revealed three major bands of molecular weights 14 Kd, 20 Kd and 26 Kd. These polypeptides were determined to have the following N-terminal sequences, respectively:

- (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 134)

(k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) and

(l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136), wherein Xaa may be any amino acid.

5 Using the assays described above, these polypeptides were shown to induce proliferation and IFN- $\gamma$  production in PBMC preparations. Figs. 1A and B show the results of such assays using PBMC preparations from a first and a second donor, respectively.

DNA sequences that encode the antigens designated as (a), (c), (d) and  
 10 (g) above were obtained by screening a genomic *M. tuberculosis* library using  $^{32}$ P end labeled degenerate oligonucleotides corresponding to the N-terminal sequence and containing *M. tuberculosis* codon bias. The screen performed using a probe corresponding to antigen (a) above identified a clone having the sequence provided in SEQ ID No. 101. The polypeptide encoded by SEQ ID No. 101 is provided in SEQ ID  
 15 No. 102. The screen performed using a probe corresponding to antigen (g) above identified a clone having the sequence provided in SEQ ID No. 52. The polypeptide encoded by SEQ ID No. 52 is provided in SEQ ID No. 53. The screen performed using a probe corresponding to antigen (d) above identified a clone having the sequence provided in SEQ ID No. 24, and the screen performed with a probe corresponding to  
 20 antigen (c) identified a clone having the sequence provided in SEQ ID No. 25.

The above amino acid sequences were compared to known amino acid sequences in the gene bank using the DNA STAR system. The database searched contains some 173,000 proteins and is a combination of the Swiss, PIR databases along with translated protein sequences (Version 87). No significant homologies to the amino  
 25 acid sequences for antigens (a)-(h) and (l) were detected.

The amino acid sequence for antigen (i) was found to be homologous to a sequence from *M. leprae*. The full length *M. leprae* sequence was amplified from genomic DNA using the sequence obtained from GENBANK. This sequence was then used to screen the *M. tuberculosis* library described below in Example 2 and a full  
 30 length copy of the *M. tuberculosis* homologue was obtained (SEQ ID No. 99).



The amino acid sequence for antigen (j) was found to be homologous to a known *M. tuberculosis* protein translated from a DNA sequence. To the best of the inventors' knowledge, this protein has not been previously shown to possess T-cell stimulatory activity. The amino acid sequence for antigen (k) was found to be related to a sequence from *M. leprae*.

In the proliferation and IFN- $\gamma$  assays described above, using three PPD positive donors, the results for representative antigens provided above are presented in Table 1:

TABLE 1

RESULTS OF PBMC PROLIFERATION AND IFN- $\gamma$  ASSAYS

Sequence	Proliferation	IFN- $\gamma$
(a)	+	-
(c)	+++	+++
(d)	++	++
(g)	+++	+++
(h)	+++	+++

In Table 1, responses that gave a stimulation index (SI) of between 2 and 4 (compared to cells cultured in medium alone) were scored as +, an SI of 4-8 or 2-4 at a concentration of 1  $\mu$ g or less was scored as ++ and an SI of greater than 8 was scored as +++. The antigen of sequence (i) was found to have a high SI (+++) for one donor and lower SI (++ and +) for the two other donors in both proliferation and IFN- $\gamma$  assays. These results indicate that these antigens are capable of inducing proliferation and/or interferon- $\gamma$  production.

EXAMPLE 2USE OF PATIENT SERA TO ISOLATE *M. TUBERCULOSIS* ANTIGENS

This example illustrates the isolation of antigens from *M. tuberculosis* lysate by screening with serum from *M. tuberculosis*-infected individuals.

Dessicated *M. tuberculosis* H37Ra (Difco Laboratories) was added to a 2% NP40 solution, and alternately homogenized and sonicated three times. The resulting suspension was centrifuged at 13,000 rpm in microfuge tubes and the supernatant put through a 0.2 micron syringe filter. The filtrate was bound to Macro Prep DEAE beads (BioRad, Hercules, CA). The beads were extensively washed with 20 mM Tris pH 7.5 and bound proteins eluted with 1M NaCl. The 1M NaCl elute was dialyzed overnight against 10 mM Tris, pH 7.5. Dialyzed solution was treated with DNase and RNase at 0.05 mg/ml for 30 min. at room temperature and then with  $\alpha$ -D-mannosidase, 0.5 U/mg at pH 4.5 for 3-4 hours at room temperature. After returning to pH 7.5, the material was fractionated via FPLC over a Bio Scale-Q-20 column (BioRad). Fractions were combined into nine pools, concentrated in a Centriprep 10 (Amicon, Beverley, MA) and then screened by Western blot for serological activity using a serum pool from *M. tuberculosis*-infected patients which was not immunoreactive with other antigens of the present invention.

The most reactive fraction was run in SDS-PAGE and transferred to PVDF. A band at approximately 85 Kd was cut out yielding the sequence:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137), wherein Xaa may be any amino acid.

Comparison of this sequence with those in the gene bank as described above, revealed no significant homologies to known sequences.

A DNA sequence that encodes the antigen designated as (m) above was obtained by screening a genomic *M. tuberculosis* Erdman strain library using labeled degenerate oligonucleotides corresponding to the N-terminal sequence of SEQ ID NO:137. A clone was identified having the DNA sequence provided in SEQ ID NO:

203. This sequence was found to encode the amino acid sequence provided in SEQ ID NO: 204. Comparison of these sequences with those in the genebank revealed some similarity to sequences previously identified in *M. tuberculosis* and *M. bovis*.

5

### EXAMPLE 3

#### PREPARATION OF DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

This example illustrates the preparation of DNA sequences encoding *M. tuberculosis* antigens by screening a *M. tuberculosis* expression library with sera  
10 obtained from patients infected with *M. tuberculosis*, or with anti-sera raised against soluble *M. tuberculosis* antigens.

#### A. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTI-SERA RAISED AGAINST *M. TUBERCULOSIS* SUPERNATANT

15 Genomic DNA was isolated from the *M. tuberculosis* strain H37Ra. The DNA was randomly sheared and used to construct an expression library using the Lambda ZAP expression system (Stratagene, La Jolla, CA). Rabbit anti-sera was generated against secretory proteins of the *M. tuberculosis* strains H37Ra, H37Rv and Erdman by immunizing a rabbit with concentrated supernatant of the *M. tuberculosis*  
20 cultures. Specifically, the rabbit was first immunized subcutaneously with 200 µg of protein antigen in a total volume of 2 ml containing 10 µg muramyl dipeptide (Calbiochem, La Jolla, CA) and 1 ml of incomplete Freund's adjuvant. Four weeks later the rabbit was boosted subcutaneously with 100 µg antigen in incomplete Freund's adjuvant. Finally, the rabbit was immunized intravenously four weeks later with 50 µg  
25 protein antigen. The anti-sera were used to screen the expression library as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

30 Thirty two clones were purified. Of these, 25 represent sequences that have not been previously identified in human *M. tuberculosis*. Recombinant antigens

10084843.022502

were expressed and purified antigens used in the immunological analysis described in Example 1. Proteins were induced by IPTG and purified by gel elution, as described in Skeiky et al., *J. Exp. Med.* 181:1527-1537, 1995. Representative sequences of DNA molecules identified in this screen are provided in SEQ ID Nos.: 1-25. The  
5 corresponding predicted amino acid sequences are shown in SEQ ID Nos. 63-87.

On comparison of these sequences with known sequences in the gene bank using the databases described above, it was found that the clones referred to hereinafter as TbRA2A, TbRA16, TbRA18, and TbRA29 (SEQ ID Nos. 76, 68, 70, 75) show some homology to sequences previously identified in *Mycobacterium leprae* but  
10 not in *M. tuberculosis*. TbRA2A was found to be a lipoprotein, with a six residue lipidation sequence being located adjacent to a hydrophobic secretory sequence. TbRA11, TbRA26, TbRA28 and TbDPEP (SEQ ID Nos.: 65, 73, 74, 53) have been previously identified in *M. tuberculosis*. No significant homologies were found to TbRA1, TbRA3, TbRA4, TbRA9, TbRA10, TbRA13, TbRA17, TbRa19, TbRA29,  
15 TbRA32, TbRA36 and the overlapping clones TbRA35 and TbRA12 (SEQ ID Nos. 63, 77, 81, 82, 64, 67, 69, 71, 75, 78, 80, 79, 66). The clone TbRa24 is overlapping with clone TbRa29.

The results of PBMC proliferation and interferon- $\gamma$  assays performed on representative recombinant antigens, and using T-cell preparations from several  
20 different *M. tuberculosis*-immune patients, are presented in Tables 2 and 3, respectively.

2052220" CH81800T  
10034843 022502

**TABLE 2**  
**RESULTS OF PBMC PROLIFERATION TO REPRESENTATIVE SOLUBLE ANTIGENS**

Antigen	Patient												
	1	2	3	4	5	6	7	8	9	10	11	12	13
TbRa1	-	-	±	++	-	-	±	±	-	-	+	±	-
TbRa3	-	±	++	-	±	-	-	++	±	-	-	-	-
TbRa9	-	-	nt	nt	++	++	nt	nt	nt	nt	nt	nt	nt
TbRa10	-	-	±	±	±	+	nt	±	-	+	±	±	-
TbRa11	±	±	+	++	++	+	nt	-	++	++	++	±	nt
TbRa12	-	-	+	+	±	++	+	±	±	-	+	-	-
TbRa16	nt	nt	nt	nt	-	+	nt	nt	nt	nt	nt	nt	nt
TbRa24	nt	nt	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
TbRa26	-	+	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
TbRa29	nt	nt	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
TbRa35	++	nt	++	++	++	++	nt	++	++	++	++	++	nt
TbRaB	nt	nt	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
TbRaC	nt	nt	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
TbRaD	nt	nt	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
AAMK	-	-	±	-	-	-	nt	-	-	-	nt	±	nt
YY	-	-	-	-	-	-	nt	-	-	-	nt	+	nt
DPEP	-	+	-	++	-	-	nt	++	±	+	±	±	nt
Control	-	-	-	-	-	-	-	-	-	-	-	-	-

nt = not tested

**TABLE 3**  
**RESULTS OF PBMC INTERFERON- $\gamma$  PRODUCTION TO REPRESENTATIVE SOLUBLE ANTIGENS**

[illegible]

In Tables 2 and 3, responses that gave a stimulation index (SI) of between 1.2 and 2 (compared to cells cultured in medium alone) were scored as  $\pm$ , a SI of 2-4 was scored as +, as SI of 4-8 or 2-4 at a concentration of 1  $\mu$ g or less was scored as ++ and an SI of greater than 8 was scored as +++. In addition, the effect of concentration on proliferation and interferon- $\gamma$  production is shown for two of the above antigens in the attached Figure. For both proliferation and interferon- $\gamma$  production, TbRa3 was scored as ++ and TbRa9 as +.

These results indicate that these soluble antigens can induce proliferation and/or interferon- $\gamma$  production in T-cells derived from an *M. tuberculosis*-immune individual.

B. USE OF SERA FROM PATIENTS HAVING PULMONARY OR PLEURAL TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

The genomic DNA library described above, and an additional H37Rv library, were screened using pools of sera obtained from patients with active tuberculosis. To prepare the H37Rv library, *M. tuberculosis* strain H37Rv genomic DNA was isolated, subjected to partial *Sau*3A digestion and used to construct an expression library using the Lambda Zap expression system (Stratagene, La Jolla, Ca). Three different pools of sera, each containing sera obtained from three individuals with active pulmonary or pleural disease, were used in the expression screening. The pools were designated TbL, TbM and TbH, referring to relative reactivity with H37Ra lysate (*i.e.*, TbL = low reactivity, TbM = medium reactivity and TbH = high reactivity) in both ELISA and immunoblot format. A fourth pool of sera from seven patients with active pulmonary tuberculosis was also employed. All of the sera lacked increased reactivity with the recombinant 38 kD *M. tuberculosis* H37Ra phosphate-binding protein.

All pools were pre-adsorbed with *E. coli* lysate and used to screen the H37Ra and H37Rv expression libraries, as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

10084843-022502

Thirty two clones were purified. Of these, 31 represented sequences that had not been previously identified in human *M. tuberculosis*. Representative sequences of the DNA molecules identified are provided in SEQ ID Nos.: 26-51 and 105. Of these, TbH-8-2 (SEQ. ID NO. 105) is a partial clone of TbH-8, and TbH-4 (SEQ. ID NO. 43) and TbH-4-FWD (SEQ. ID NO. 44) are non-contiguous sequences from the same clone. Amino acid sequences for the antigens hereinafter identified as Tb38-1, TbH-4, TbH-8, TbH-9, and TbH-12 are shown in SEQ ID Nos.: 88-92. Comparison of these sequences with known sequences in the gene bank using the databases identified above revealed no significant homologies to TbH-4, TbH-8, TbH-9 and TbM-3, although weak homologies were found to TbH-9. TbH-12 was found to be homologous to a 34 kD antigenic protein previously identified in *M. paratuberculosis* (Acc. No. S28515). Tb38-1 was found to be located 34 base pairs upstream of the open reading frame for the antigen ESAT-6 previously identified in *M. bovis* (Acc. No. U34848) and in *M. tuberculosis* (Sorensen et al., *Infect. Immun.* 63:1710-1717, 1995).

Probes derived from Tb38-1 and TbH-9, both isolated from an H37Ra library, were used to identify clones in an H37Rv library. Tb38-1 hybridized to Tb38-1F2, Tb38-1F3, Tb38-1F5 and Tb38-1F6 (SEQ. ID NOS. 112, 113, 116, 118, and 119). (SEQ ID NOS. 112 and 113 are non-contiguous sequences from clone Tb38-1F2.) Two open reading frames were deduced in Tb38-1F2; one corresponds to Tb37FL (SEQ. ID. NO. 114), the second, a partial sequence, may be the homologue of Tb38-1 and is called Tb38-IN (SEQ. ID NO. 115). The deduced amino acid sequence of Tb38-1F3 is presented in SEQ. ID. NO. 117. A TbH-9 probe identified three clones in the H37Rv library: TbH-9-FL (SEQ. ID NO. 106), which may be the homologue of TbH-9 (R37Ra), TbH-9-1 (SEQ. ID NO. 108), and TbH-9-4 (SEQ. ID NO. 110), all of which are highly related sequences to TbH-9. The deduced amino acid sequences for these three clones are presented in SEQ ID NOS. 107, 109 and 111.

Further screening of the *M. tuberculosis* genomic DNA library, as described above, resulted in the recovery of ten additional reactive clones, representing seven different genes. One of these genes was identified as the 38 Kd antigen discussed



above, one was determined to be identical to the 14Kd alpha crystallin heat shock protein previously shown to be present in *M. tuberculosis*, and a third was determined to be identical to the antigen TbH-8 described above. The determined DNA sequences for the remaining five clones (hereinafter referred to as TbH-29, TbH-30, TbH-32 and 5 TbH-33) are provided in SEQ ID NO: 138-141, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 142-145, respectively. The DNA and amino acid sequences for these antigens were compared with those in the gene bank as described above. No homologies were found to the 5' end of TbH-29 (which contains the reactive open reading frame), although the 3' end of TbH-29 was 10 found to be identical to the *M. tuberculosis* cosmid Y227. TbH-32 and TbH-33 were found to be identical to the previously identified *M. tuberculosis* insertion element IS6110 and to the *M. tuberculosis* cosmid Y50, respectively. No significant homologies to TbH-30 were found.

Positive phagemid from this additional screening were used to infect *E.* 15 *coli* XL-1 Blue MRF', as described in Sambrook et al., *supra*. Induction of recombinant protein was accomplished by the addition of IPTG. Induced and uninduced lysates were run in duplicate on SDS-PAGE and transferred to nitrocellulose filters. Filters were reacted with human *M. tuberculosis* sera (1:200 dilution) reactive with TbH and a rabbit sera (1:200 or 1:250 dilution) reactive with the N-terminal 4 Kd portion of lacZ. 20 Sera incubations were performed for 2 hours at room temperature. Bound antibody was detected by addition of <sup>125</sup>I-labeled Protein A and subsequent exposure to film for variable times ranging from 16 hours to 11 days. The results of the immunoblots are summarized in Table 4.

TABLE 4

<u>Antigen</u>	<u>Human M. tb Sera</u>	<u>Anti-lacZ Sera</u>
TbH-29	45 Kd	45 Kd
TbH-30	No reactivity	29 Kd
TbH-32	12 Kd	12 Kd
TbH-33	16 Kd	16 Kd

Positive reaction of the recombinant human *M. tuberculosis* antigens with both the human *M. tuberculosis* sera and anti-lacZ sera indicate that reactivity of the human *M. tuberculosis* sera is directed towards the fusion protein. Antigens reactive with the anti-lacZ sera but not with the human *M. tuberculosis* sera may be the result of the human *M. tuberculosis* sera recognizing conformational epitopes, or the antigen-antibody binding kinetics may be such that the 2 hour sera exposure in the immunoblot is not sufficient.

The results of T-cell assays performed on Tb38-1, ESAT-6 and other representative recombinant antigens are presented in Tables 5A, B and 6, respectively, below:

TABLE 5A  
RESULTS OF PBMC PROLIFERATION TO REPRESENTATIVE ANTIGENS

Antigen	Donor										
	1	2	3	4	5	6	7	8	9	10	11
Tb38.1	+++	+	-	-	-	++	-	+	-	++	+++
ESAT-6	+++	+	+	+	-	+	-	+	+	++	+++
TbH-9	++	++	-	++	±	±	++	++	++	++	++

**TABLE 5B**  
**RESULTS OF PBMC INTERFERON- $\gamma$  PRODUCTION TO REPRESENTATIVE ANTIGENS**

Antigen	Donor										
	1	2	3	4	5	6	7	8	9	10	11
Tb38.1	+++	+	-	+	+	+++	-	++	-	+++	+++
ESAT-6	+++	+	+	+	+-	+	-	+	+	+++	+++
TbH-9	++	++	-	+++	$\pm$	$\pm$	+++	+++	++	+++	++

5

**TABLE 6**  
**SUMMARY OF T-CELL RESPONSES TO REPRESENTATIVE ANTIGENS**

Antigen	Proliferation			Interferon- $\gamma$			total
	patient 4	patient 5	patient 6	patient 4	patient 5	patient 6	
TbH9	++	++	++	+++	++	++	13
TbM7	-	+	-	++	+	-	4
TbH5	-	+	+	++	++	++	8
TbL23	-	+	$\pm$	++	++	+	7.5
TbH4	-	++	$\pm$	++	++	$\pm$	7
- control	-	-	-	-	-	-	0

10                These results indicate that both the inventive *M. tuberculosis* antigens and ESAT-6 can induce proliferation and/or interferon- $\gamma$  production in T-cells derived from an *M. tuberculosis*-immune individual. To the best of the inventors' knowledge, ESAT-6 has not been previously shown to stimulate human immune responses

15                A set of six overlapping peptides covering the amino acid sequence of the antigen Tb38-1 was constructed using the method described in Example 6. The sequences of these peptides, hereinafter referred to as pep1-6, are provided in SEQ ID Nos. 93-98, respectively. The results of T-cell assays using these peptides are shown in Tables 7 and 8. These results confirm the existence, and help to localize T-cell epitopes within Tb38-1 capable of inducing proliferation and interferon- $\gamma$  production in T-cells  
20                derived from an *M. tuberculosis* immune individual.

2052220 " E4B4800T



**TABLE 8**

[illegible]

Studies were undertaken to determine whether the antigens TbH-9 and Tb38-1 represent cellular proteins or are secreted into *M. tuberculosis* culture media. In the first study, rabbit sera were raised against A) secretory proteins of *M. tuberculosis*, B) the known secretory recombinant *M. tuberculosis* antigen 85b, C) recombinant Tb38-1 and D) recombinant TbH-9, using protocols substantially the same as that as described in Example 3A. Total *M. tuberculosis* lysate, concentrated supernatant of *M. tuberculosis* cultures and the recombinant antigens 85b, TbH-9 and Tb38-1 were resolved on denaturing gels, immobilized on nitrocellulose membranes and duplicate blots were probed using the rabbit sera described above.

The results of this analysis using control sera (panel I) and antisera (panel II) against secretory proteins, recombinant 85b, recombinant Tb38-1 and recombinant TbH-9 are shown in Figures 3A-D, respectively, wherein the lane designations are as follows: 1) molecular weight protein standards; 2) 5 µg of *M. tuberculosis* lysate; 3) 5 µg secretory proteins; 4) 50 ng recombinant Tb38-1; 5) 50 ng recombinant TbH-9; and 6) 50 ng recombinant 85b. The recombinant antigens were engineered with six terminal histidine residues and would therefore be expected to migrate with a mobility approximately 1 kD larger than the native protein. In Figure 3D, recombinant TbH-9 is lacking approximately 10 kD of the full-length 42 kD antigen, hence the significant difference in the size of the immunoreactive native TbH-9 antigen in the lysate lane (indicated by an arrow). These results demonstrate that Tb38-1 and TbH-9 are intracellular antigens and are not actively secreted by *M. tuberculosis*.

The finding that TbH-9 is an intracellular antigen was confirmed by determining the reactivity of TbH-9-specific human T cell clones to recombinant TbH-9, secretory *M. tuberculosis* proteins and PPD. A TbH-9-specific T cell clone (designated 131TbH-9) was generated from PBMC of a healthy PPD-positive donor. The proliferative response of 131TbH-9 to secretory proteins, recombinant TbH-9 and a control *M. tuberculosis* antigen, TbRa11, was determined by measuring uptake of tritiated thymidine, as described in Example 1. As shown in Figure 4A, the clone 131TbH-9 responds specifically to TbH-9, showing that TbH-9 is not a significant component of *M. tuberculosis* secretory proteins. Figure 4B shows the production of IFN-γ by a second TbH-9-specific T cell clone

10084843 " 022502  
205220 " 5488001

(designated PPD 800-10) prepared from PBMC from a healthy PPD-positive donor, following stimulation of the T cell clone with secretory proteins, PPD or recombinant TbH-9. These results further confirm that TbH-9 is not secreted by *M. tuberculosis*.

C. USE OF SERA FROM PATIENTS HAVING EXTRAPULMONARY TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). The resulting library was screened using pools of sera obtained from individuals with extrapulmonary tuberculosis, as described above in Example 3B, with the secondary antibody being goat anti-human IgG + A + M (H+L) conjugated with alkaline phosphatase.

Eighteen clones were purified. Of these, 4 clones (hereinafter referred to as XP14, XP24, XP31 and XP32) were found to bear some similarity to known sequences. The determined DNA sequences for XP14, XP24 and XP31 are provided in SEQ ID Nos.: 156-158, respectively, with the 5' and 3' DNA sequences for XP32 being provided in SEQ ID Nos.: 159 and 160, respectively. The predicted amino acid sequence for XP14 is provided in SEQ ID No: 161. The reverse complement of XP14 was found to encode the amino acid sequence provided in SEQ ID No.: 162.

Comparison of the sequences for the remaining 14 clones (hereinafter referred to as XP1-XP6, XP17-XP19, XP22, XP25, XP27, XP30 and XP36) with those in the genebank as described above, revealed no homologies with the exception of the 3' ends of XP2 and XP6 which were found to bear some homology to known *M. tuberculosis* cosmids. The DNA sequences for XP27 and XP36 are shown in SEQ ID Nos.: 163 and 164, respectively, with the 5' sequences for XP4, XP5, XP17 and XP30 being shown in SEQ ID Nos: 165-168, respectively, and the 5' and 3' sequences for XP2, XP3, XP6, XP18, XP19, XP22 and XP25 being shown in SEQ ID Nos: 169 and 170; 171 and 172; 173 and 174; 175 and 176; 177 and 178; 179 and 180; and 181 and 182, respectively. XP1 was found to overlap with the DNA sequences for TbH4, disclosed above. The full-length DNA sequence for TbH4-XP1 is provided in SEQ ID No.: 183. This DNA sequence was found to contain an

10084843-022502

open reading frame encoding the amino acid sequence shown in SEQ ID No: 184. The reverse complement of TbH4-XP1 was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID No.: 185. The DNA sequence for XP36 was found to contain two open reading frames encoding the amino acid sequence shown in SEQ ID Nos.: 186 and 187, with the reverse complement containing an open reading frame encoding the amino acid sequence shown in SEQ ID No.: 188.

Recombinant XP1 protein was prepared as described above in Example 3B, with a metal ion affinity chromatography column being employed for purification. As illustrated in Figures 8A-B and 9A-B, using the assays described herein, recombinant XP1 was found to stimulate cell proliferation and IFN- $\gamma$  production in T cells isolated from an *M. tuberculosis*-immune donors.

D. USE OF A LYSATE POSITIVE SERUM POOL FROM PATIENTS HAVING TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda Screen expression system (Novagen, Madison, WI), as described below in Example 6. Pooled serum obtained from *M. tuberculosis*-infected patients and that was shown to react with *M. tuberculosis* lysate but not with the previously expressed proteins 38kD, Tb38-1, TbRa3, TbH4, DPEP and TbRa11, was used to screen the expression library as described above in Example 3B, with the secondary antibody being goat anti-human IgG + A + M (H+L) conjugated with alkaline phosphatase.

Twenty-seven clones were purified. Comparison of the determined cDNA sequences for these clones revealed no significant homologies to 10 of the clones (hereinafter referred to as LSER-10, LSER-11, LSER-12, LSER-13, LSER-16, LSER-18, LSER-23, LSER-24, LSER-25 and LSER-27). The determined 5' cDNA sequences for LSER-10, LSER-11, LSER-12, LSER-13, LSER-16 and LSER-25 are provided in SEQ ID NO: 242-247, respectively, with the corresponding predicted amino acid sequences for LSER-10, LSER-12, LSER-13, LSER-16 and LSER-25 being provided in SEQ ID NO: 248-252, respectively. The determined full-length cDNA sequences for LSER-18, LSER-23, LSER-24



and LSER-27 are shown in SEQ ID NO: 253-256, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 257-260. The remaining seventeen clones were found to show similarities to unknown sequences previously identified in *M. tuberculosis*. The determined 5' cDNA sequences for sixteen of these clones (hereinafter referred to as LSER-1, LSER-3, LSER-4, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30) are provided in SEQ ID NO: 261-276, respectively, with the corresponding predicted amino acid sequences for LSER-1, LSER-3, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30 being provided in SEQ ID NO: 277-291, respectively. The determined full-length cDNA sequence for the clone LSER-9 is provided in SEQ ID NO: 292. The reverse complement of LSER-6 (SEQ ID NO: 293) was found to encode the predicted amino acid sequence of SEQ ID NO: 294.

E. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTI-SERA RAISED AGAINST *M. TUBERCULOSIS* FRACTIONATED PROTEINS

*M. tuberculosis* lysate was prepared as described above in Example 2. The resulting material was fractionated by HPLC and the fractions screened by Western blot for serological activity with a serum pool from *M. tuberculosis*-infected patients which showed little or no immunoreactivity with other antigens of the present invention. Rabbit anti-sera was generated against the most reactive fraction using the method described in Example 3A. The anti-sera was used to screen an *M. tuberculosis* Erdman strain genomic DNA expression library prepared as described above. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones determined.

Ten different clones were purified. Of these, one was found to be TbRa35, described above, and one was found to be the previously identified *M. tuberculosis* antigen, HSP60. Of the remaining eight clones, seven (hereinafter referred to as RDIF2, RDIF5, RDIF8, RDIF10, RDIF11 and RDIF 12) were found to bear some similarity to previously identified *M. tuberculosis* sequences. The determined DNA sequences for RDIF2, RDIF5, RDIF8, RDIF10 and RDIF11 are provided in SEQ ID Nos.: 189-193, respectively, with the

10084943.022502

corresponding predicted amino acid sequences being provided in SEQ ID Nos: 194-198, respectively. The 5' and 3' DNA sequences for RDIF12 are provided in SEQ ID Nos.: 199 and 200, respectively. No significant homologies were found to the antigen RDIF-7. The determined DNA and predicted amino acid sequences for RDIF7 are provided in SEQ ID Nos.: 201 and 202, respectively. One additional clone, referred to as RDIF6 was isolated, however, this was found to be identical to RDIF5.

Recombinant RDIF6, RDIF8, RDIF10 and RDIF11 were prepared as described above. As shown in Figures 8A-B and 9A-B, these antigens were found to stimulate cell proliferation and IFN- $\gamma$  production in T cells isolated from *M. tuberculosis*-immune donors.

#### EXAMPLE 4

##### PURIFICATION AND CHARACTERIZATION OF A POLYPEPTIDE FROM TUBERCULIN PURIFIED PROTEIN DERIVATIVE

An *M. tuberculosis* polypeptide was isolated from tuberculin purified protein derivative (PPD) as follows.

PPD was prepared as published with some modification (Seibert, F. et al., Tuberculin purified protein derivative. Preparation and analyses of a large quantity for standard. The American Review of Tuberculosis 44:9-25, 1941).

*M. tuberculosis* Rv strain was grown for 6 weeks in synthetic medium in roller bottles at 37°C. Bottles containing the bacterial growth were then heated to 100° C in water vapor for 3 hours. Cultures were sterile filtered using a 0.22  $\mu$  filter and the liquid phase was concentrated 20 times using a 3 kD cut-off membrane. Proteins were precipitated once with 50% ammonium sulfate solution and eight times with 25% ammonium sulfate solution. The resulting proteins (PPD) were fractionated by reverse phase liquid chromatography (RP-HPLC) using a C18 column (7.8 x 300 mM; Waters, Milford, MA) in a Biocad HPLC system (Perseptive Biosystems, Framingham, MA). Fractions were eluted from the column with a linear gradient from 0-100% buffer (0.1% TFA in acetonitrile). The flow rate was 10 ml/minute and eluent was monitored at 214 nm and 280 nm.

Six fractions were collected, dried, suspended in PBS and tested individually in *M. tuberculosis*-infected guinea pigs for induction of delayed type hypersensitivity (DTH) reaction. One fraction was found to induce a strong DTH reaction and was subsequently fractionated further by RP-HPLC on a microbore Vydac C18 column (Cat. No. 218TP5115) in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted with a linear gradient from 5-100% buffer (0.05% TFA in acetonitrile) with a flow rate of 80  $\mu$ l/minute. Eluent was monitored at 215 nm. Eight fractions were collected and tested for induction of DTH in *M. tuberculosis*-infected guinea pigs. One fraction was found to induce strong DTH of about 16 mm induration. The other fractions did not induce detectable DTH. The positive fraction was submitted to SDS-PAGE gel electrophoresis and found to contain a single protein band of approximately 12 kD molecular weight.

This polypeptide, herein after referred to as DPPD, was sequenced from the amino terminal using a Perkin Elmer/Applied Biosystems Division Procise 492 protein sequencer as described above and found to have the N-terminal sequence shown in SEQ ID No.: 129. Comparison of this sequence with known sequences in the gene bank as described above revealed no known homologies. Four cyanogen bromide fragments of DPPD were isolated and found to have the sequences shown in SEQ ID Nos.: 130-133. A subsequent search of the *M. tuberculosis* genome database released by the Institute for Genomic Research revealed a match of the DPPD partial amino acid sequence with a sequence present within the *M. tuberculosis* cosmid MTY21C12. An open reading frame of 336 bp was identified. The full-length DNA sequence for DPPD is provided in SEQ ID NO: 240, with the corresponding full-length amino acid sequence being provided in SEQ ID NO: 241.

The ability of the antigen DPPD to stimulate human PBMC to proliferate and to produce IFN- $\gamma$  was assayed as described in Example 1. As shown in Table 9, DPPD was found to stimulate proliferation and elicit production of large quantities of IFN- $\gamma$ ; more than that elicited by commercial PPD.

TABLE 9  
RESULTS OF PROLIFERATION AND INTERFERON- $\gamma$  ASSAYS TO DPPD

PBMC Donor	Stimulator	Proliferation (CPM)	IFN- $\gamma$ (OD <sub>450</sub> )
A	Medium	1,089	0.17
	PPD (commercial)	8,394	1.29
	DPPD	13,451	2.21
B	Medium	450	0.09
	PPD (commercial)	3,929	1.26
	DPPD	6,184	1.49
C	Medium	541	0.11
	PPD (commercial)	8,907	0.76
	DPPD	23,024	>2.70

EXAMPLE 5

USE OF SERA FROM TUBERCULOSIS-INFECTED MONKEYS TO IDENTIFY DNA SEQUENCES  
ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). Serum samples were obtained from a cynomolgous monkey 18, 33, 51 and 56 days following infection with *M. tuberculosis* Erdman strain. These samples were pooled and used to screen the *M. tuberculosis* genomic DNA expression library using the procedure described above in Example 3C.

Twenty clones were purified. The determined 5' DNA sequences for the clones referred to as MO-1, MO-2, MO-4, MO-8, MO-9, MO-26, MO-28, MO-29, MO-30, MO-34 and MO-35 are provided in SEQ ID NO: 215-225, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 226-236. The full-length DNA sequence of the clone MO-10 is provided in SEQ ID NO: 237, with the corresponding

205220" 24848007

predicted amino acid sequence being provided in SEQ ID NO: 238. The 3' DNA sequence for the clone MO-27 is provided in SEQ ID NO: 239.

Clones MO-1, MO-30 and MO-35 were found to show a high degree of relatedness and showed some homology to a previously identified unknown *M. tuberculosis* sequence and to cosmid MTCI237. MO-2 was found to show some homology to aspartokinase from *M. tuberculosis*. Clones MO-3, MO-7 and MO-27 were found to be identical and to show a high degree of relatedness to MO-5. All four of these clones showed some homology to *M. tuberculosis* heat shock protein 70. MO-27 was found to show some homology to *M. tuberculosis* cosmid MTCY339. MO-4 and MO-34 were found to show some homology to cosmid SCY21B4 and *M. smegmatis* integration host factor, and were both found to show some homology to a previously identified, unknown *M. tuberculosis* sequence. MO-6 was found to show some homology to *M. tuberculosis* heat shock protein 65. MO-8, MO-9, MO-10, MO-26 and MO-29 were found to be highly related to each other and to show some homology to *M. tuberculosis* dihydrolipamide succinyltransferase. MO-28, MO-31 and MO-32 were found to be identical and to show some homology to a previously identified *M. tuberculosis* protein. MO-33 was found to show some homology to a previously identified 14 kDa *M. tuberculosis* heat shock protein.

Further studies using the above protocol resulted in the isolation of an additional four clones, hereinafter referred to as MO-12, MO-13, MO-19 and MO-39. The determined 5' cDNA sequences for these clones are provided in SEQ ID NO: 295-298, respectively, with the corresponding predicted protein sequences being provided in SEQ ID NO: 299-302, respectively. Comparison of these sequences with those in the gene bank as described above revealed no significant homologies to MO-39. MO-12, MO-13 and MO-19 were found to show some homologies to unknown sequences previously isolated from *M. tuberculosis*.

#### EXAMPLE 6

#### ISOLATION OF DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS BY SCREENING OF A NOVEL EXPRESSION LIBRARY

10084843.022502  
205220" CH8800T

This example illustrates isolation of DNA sequences encoding *M. tuberculosis* antigens by screening of a novel expression library with sera from *M. tuberculosis*-infected patients that were shown to be unreactive with a panel of the recombinant *M. tuberculosis* antigens TbRa11, TbRa3, Tb38-1, TbH4, TbF and 38 kD.

Genomic DNA from *M. tuberculosis* Erdman strain was randomly sheared to an average size of 2 kb, and blunt ended with Klenow polymerase, followed by the addition of EcoRI adaptors. The insert was subsequently ligated into the Screen phage vector (Novagen, Madison, WI) and packaged *in vitro* using the PhageMaker extract (Novagen). The resulting library was screened with sera from several *M. tuberculosis* donors that had been shown to be negative on a panel of previously identified *M. tuberculosis* antigens as described above in Example 3B.

A total of 22 different clones were isolated. By comparison, screening of the  $\lambda$  Zap library described above using the same sera did not result in any positive hits. One of the clones was found to represent TbRa11, described above. The determined 5' cDNA sequences for 19 of the remaining 21 clones (hereinafter referred to as Erdsn1, Erdsn2, Erdsn4-Erdsn10, Erdsn12-18, Erdsn21-Erdsn23 and Erdsn25) are provided in SEQ ID NO: 303-322, respectively, with the determined 3' cDNA sequences for Erdsn1, Erdsn2, Erdsn4, Erdsn-5, Erdsn-7-Erdsn10, Erdsn12-18, Erdsn21-Erdsn23 and Erdsn25 being provided in SEQ ID NO: 323-341, respectively. The complete cDNA insert sequence for the clone Erdsn24 is provided in SEQ ID NO: 342. Comparison of the determined cDNA sequences with those in the gene bank revealed no significant homologies to the sequences provided in SEQ ID NO: 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341. The sequences of SEQ ID NO: 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342 were found to show some homology to unknown sequences previously identified in *M. tuberculosis*.

#### EXAMPLE 7

##### ISOLATION OF SOLUBLE M. TUBERCULOSIS ANTIGENS

##### USING MASS SPECTROMETRY

10084843.022502

This example illustrates the use of mass spectrometry to identify soluble *M. tuberculosis* antigens.

In a first approach, *M. tuberculosis* culture filtrate was screened by Western analysis using serum from a tuberculosis-infected individual. The reactive bands were excised from a silver stained gel and the amino acid sequences determined by mass spectrometry. The determined amino acid sequence for one of the isolated antigens is provided in SEQ ID NO: 343. Comparison of this sequence with those in the gene bank revealed homology to the 85b precursor antigen previously identified in *M. tuberculosis*.

In a second approach, the high molecular weight region of *M. tuberculosis* culture supernatant was studied. This area may contain immunodominant antigens which may be useful in the diagnosis of *M. tuberculosis* infection. Two known monoclonal antibodies, IT42 and IT57 (available from the Center for Disease Control, Atlanta, GA), show reactivity by Western analysis to antigens in this vicinity, although the identity of the antigens remains unknown. In addition, unknown high-molecular weight proteins have been described as containing a surrogate marker for *M. tuberculosis* infection in HIV-positive individuals (*Jnl. Infect. Dis.*, 176:133-143, 1997). To determine the identity of these antigens, two-dimensional gel electrophoresis and two-dimensional Western analysis were performed using the antibodies IT57 and IT42. Five protein spots in the high molecular weight region were identified, individually excised, enzymatically digested and subjected to mass spectrometric analysis.

The determined amino acid sequences for three of these spots (referred to as spots 1, 2 and 4) are provided in SEQ ID NO: 344, 345-346 and 347, respectively. Comparison of these sequences with those in the gene bank revealed that spot 1 is the previously identified PcK-1, a phosphoenolpyruvate kinase. The two sequences isolated from spot 2 were determined to be from two DNAs, previously identified in *M. tuberculosis* as heat shock proteins. Spot 4 was determined to be the previously identified *M. tuberculosis* protein Kat G. To the best of the inventors' knowledge, neither PcK-1 nor the two DNAs have previously been shown to have utility in the diagnosis of *M. tuberculosis* infection.

#### EXAMPLE 8

205220-248001

### USE OF REPRESENTATIVE ANTIGENS FOR DIAGNOSIS OF TUBERCULOSIS

This example illustrates the effectiveness of several representative polypeptides in skin tests for the diagnosis of *M. tuberculosis* infection.

Individuals were injected intradermally with 100  $\mu$ l of either PBS or PBS plus Tween 20<sup>TM</sup> containing either 0.1  $\mu$ g of protein (for TbH-9 and TbRa35) or 1.0  $\mu$ g of protein (for TbRa38-1). Induration was measured between 5-7 days after injection, with a response of 5 mm or greater being considered positive. Of the 20 individuals tested, 2 were PPD negative and 18 were PPD positive. Of the PPD positive individuals, 3 had active tuberculosis, 3 had been previously infected with tuberculosis and 9 were healthy. In a second study, 13 PPD positive individuals were tested with 0.1  $\mu$ g TbRa11 in either PBS or PBS plus Tween 20<sup>TM</sup> as described above. The results of both studies are shown in Table 10.

TABLE 10  
RESULTS OF DTH TESTING WITH REPRESENTATIVE ANTIGENS

	TbH-9 Pos/Total	Tb38-1 Pos/Total	TbRa35 Pos/Total	Cumulative Pos/Total	TbRa11 Pos/Total
PPD negative	0/2	0/2	0/2	0/2	
PPD positive					
healthy	5/9	4/9	4/9	6/9	1/4
prior TB	3/5	2/5	2/5	4/5	3/5
active	3/4	3/4	0/4	4/4	1/4
TOTAL	11/18	9/18	6/18	14/18	5/13

### EXAMPLE 9

#### SYNTHESIS OF SYNTHETIC POLYPEPTIDES

Polypeptides may be synthesized on a Millipore 9050 peptide synthesizer using Fmoc chemistry with HPTU (O-Benzotriazole-N,N,N',N'-tetramethyluronium

205220" CH84800T



hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0%-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray mass spectrometry and by amino acid analysis.

#### EXAMPLE 10

##### PREPARATION AND CHARACTERIZATION OF *M. TUBERCULOSIS* FUSION PROTEINS

A fusion protein containing TbRa3, the 38 kD antigen and Tb38-1 was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR in order to facilitate their fusion and the subsequent expression of the fusion protein TbRa3-38 kD-Tb38-1. TbRa3, 38 kD and Tb38-1 DNA was used to perform PCR using the primers PDM-64 and PDM-65 (SEQ ID NO: 146 and 147), PDM-57 and PDM-58 (SEQ ID NO: 148 and 149), and PDM-69 and PDM-60 (SEQ ID NO: 150 and 151), respectively. In each case, the DNA amplification was performed using 10 µl 10X Pfu buffer, 2 µl 10 mM dNTPs, 2 µl each of the PCR primers at 10 µM concentration, 81.5 µl water, 1.5 µl Pfu DNA polymerase (Stratagene, La Jolla, CA) and 1 µl DNA at either 70 ng/µl (for TbRa3) or 50 ng/µl (for 38 kD and Tb38-1). For TbRa3, denaturation at 94°C was performed for 2 min, followed by 40 cycles of 96°C for 15 sec and 72°C for 1 min, and lastly by 72°C for 4 min. For 38 kD, denaturation at 96°C was performed for 2 min, followed by 40 cycles of 96°C for 30 sec, 68°C for 15 sec and 72°C for 3 min, and finally by 72°C for 4 min. For Tb38-1 denaturation at 94°C for 2 min was followed by 10 cycles of 96°C for 15 sec, 68°C for 15 sec and 72°C for

10084843 022502

1.5 min, 30 cycles of 96°C for 15 sec, 64°C for 15 sec and 72°C for 1.5, and finally by 72°C for 4 min.

The TbRa3 PCR fragment was digested with NdeI and EcoRI and cloned directly into pT7<sup>+</sup>L2 IL 1 vector using NdeI and EcoRI sites. The 38 kD PCR fragment was digested with Sse8387I, treated with T4 DNA polymerase to make blunt ends and then digested with EcoRI for direct cloning into the pT7<sup>+</sup>L2Ra3-1 vector which was digested with StuI and EcoRI. The 38-1 PCR fragment was digested with Eco47III and EcoRI and directly subcloned into pT7<sup>+</sup>L2Ra3/38kD-17 digested with the same enzymes. The whole fusion was then transferred to pET28b – using NdeI and EcoRI sites. The fusion construct was confirmed by DNA sequencing.

The expression construct was transformed into BLR pLys S *E. coli* (Novagen, Madison, WI) and grown overnight in LB broth with kanamycin (30 µg/ml) and chloramphenicol (34 µg/ml). This culture (12 ml) was used to inoculate 500 ml 2XYT with the same antibiotics and the culture was induced with IPTG at an OD<sub>560</sub> of 0.44 to a final concentration of 1.2 mM. Four hours post-induction, the bacteria were harvested and sonicated in 20 mM Tris (8.0), 100 mM NaCl, 0.1% DOC, 20 µg/ml Leupeptin, 20 mM PMSF followed by centrifugation at 26,000 X g. The resulting pellet was resuspended in 8 M urea, 20 mM Tris (8.0), 100 mM NaCl and bound to Pro-bond nickel resin (Invitrogen, Carlsbad, CA). The column was washed several times with the above buffer then eluted with an imidazole gradient (50 mM, 100 mM, 500 mM imidazole was added to 8 M urea, 20 mM Tris (8.0), 100 mM NaCl). The eluates containing the protein of interest were then dialyzed against 10 mM Tris (8.0).

The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbRa3-38 kD-Tb38-1) are provided in SEQ ID NO: 152 and 153, respectively.

A fusion protein containing the two antigens TbH-9 and Tb38-1 (hereinafter referred to as TbH9-Tb38-1) without a hinge sequence, was prepared using a similar procedure to that described above. The DNA sequence for the TbH9-Tb38-1 fusion protein is provided in SEQ ID NO: 156.

The ability of the fusion protein TbH9-Tb38-1 to induce T cell proliferation and IFN- $\gamma$  production in PBMC preparations was examined using the protocol described above in Example 1. PBMC from three donors were employed: one who had been previously shown to respond to TbH9 but not Tb38-1 (donor 131); one who had been shown to respond to Tb38-1 but not TbH9 (donor 184); and one who had been shown to respond to both antigens (donor 201). The results of these studies (Figs. 5-7, respectively) demonstrate the functional activity of both the antigens in the fusion protein.

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and DPEP was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR and cloned into vectors essentially as described above, with the primers PDM-69 (SEQ ID NO:150 and PDM-83 (SEQ ID NO: 205) being used for amplification of the Tb38-1A fragment. Tb38-1A differs from Tb38-1 by a DraI site at the 3' end of the coding region that keeps the final amino acid intact while creating a blunt restriction site that is in frame. The TbRa3/38kD/Tb38-1A fusion was then transferred to pET28b using NdeI and EcoRI sites.

DPEP DNA was used to perform PCR using the primers PDM-84 and PDM-85 (SEQ ID NO: 206 and 207, respectively) and 1  $\mu$ l DNA at 50 ng/ $\mu$ l. Denaturation at 94 °C was performed for 2 min, followed by 10 cycles of 96 °C for 15 sec, 68 °C for 15 sec and 72 °C for 1.5 min; 30 cycles of 96 °C for 15 sec, 64 °C for 15 sec and 72 °C for 1.5 min; and finally by 72 °C for 4 min. The DPEP PCR fragment was digested with EcoRI and Eco72I and clones directly into the pET28Ra3/38kD/38-1A construct which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-2) are provided in SEQ ID NO: 208 and 209, respectively.

The reactivity of the fusion protein TbF-2 with sera from *M. tuberculosis*-infected patients was examined by ELISA using the protocol described above. The results of these studies (Table 11) demonstrate that all four antigens function independently in the fusion protein.

10084843.022502

TABLE 11

REACTIVITY OF TbF-2 FUSION RECOMBINANT WITH TB AND NORMAL SERA

Serum ID	Status	TbF OD450	Status	TbF-2 OD450	Status	ELISA Reactivity			
						38 kD	TbRa3	Tb38-1	DPEP
B931-40	TB	0.57	+	0.321	+	-	+	-	+
B931-41	TB	0.601	+	0.396	+	+	+	+	-
B931-109	TB	0.494	+	0.404	+	+	+	±	-
B931-132	TB	1.502	+	1.292	+	+	+	+	±
5004	TB	1.806	+	1.666	+	±	±	+	-
15004	TB	2.862	+	2.468	+	+	+	+	-
39004	TB	2.443	+	1.722	+	+	+	+	-
68004	TB	2.871	+	2.575	+	+	+	+	-
99004	TB	0.691	+	0.971	+	-	±	+	-
107004	TB	0.875	+	0.732	+	-	±	+	-
92004	TB	1.632	+	1.394	+	+	±	±	-
97004	TB	1.491	+	1.979	+	+	±	-	+
118004	TB	3.182	+	3.045	+	+	±	-	-
173004	TB	3.644	+	3.578	+	+	+	+	-
175004	TB	3.332	+	2.916	+	+	+	-	-
274004	TB	3.696	+	3.716	+	-	+	-	+
276004	TB	3.243	+	2.56	+	-	-	+	-
282004	TB	1.249	+	1.234	+	+	-	-	-
289004	TB	1.373	+	1.17	+	-	+	-	-
308004	TB	3.708	+	3.355	+	-	-	+	-
314004	TB	1.663	+	1.399	+	-	-	+	-
317004	TB	1.163	+	0.92	+	+	-	-	-
312004	TB	1.709	+	1.453	+	-	+	-	-
380004	TB	0.238	-	0.461	+	-	±	-	+
451004	TB	0.18	-	0.2	-	-	-	-	±
478004	TB	0.188	-	0.469	+	-	-	-	±
410004	TB	0.384	+	2.392	+	±	-	-	+
411004	TB	0.306	+	0.874	+	-	+	-	+
421004	TB	0.357	+	1.456	+	-	+	-	+
528004	TB	0.047	-	0.196	-	-	-	-	+
A6-87	Normal	0.094	-	0.063	-	-	-	-	-
A6-88	Normal	0.214	-	0.19	-	-	-	-	-
A6-89	Normal	0.248	-	0.125	-	-	-	-	-
A6-90	Normal	0.179	-	0.206	-	-	-	-	-
A6-91	Normal	0.135	-	0.151	-	-	-	-	-
A6-92	Normal	0.064	-	0.097	-	-	-	-	-
A6-93	Normal	0.072	-	0.098	-	-	-	-	-
A6-94	Normal	0.072	-	0.064	-	-	-	-	-
A6-95	Normal	0.125	-	0.159	-	-	-	-	-
A6-96	Normal	0.121	-	0.12	-	-	-	-	-
Cut-off		0.284		0.266					

20220420 10:43:43

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and TbH4 was prepared as follows.

Genomic *M. tuberculosis* DNA was used to PCR full-length TbH4 (FL TbH4) with the primers PDM-157 and PDM-160 (SEQ ID NO: 348 and 349, respectively) and 2 µl DNA at 100 ng/µl. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 61 °C for 20 sec and 72 °C for 5 min; and finally by annealing at 72 °C for 10 min. The FL TbH4 PCR fragment was digested with EcoRI and Sca I (New England Biolabs.) and cloned directly into the pET28Ra3/38kD/38-1A construct described above which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-6) are provided in SEQ ID NO: 350 and 351, respectively.

A fusion protein containing the antigen 38kD and DPEP separated by a linker was prepared as follows.

38 kD DNA was used to perform PCR using the primers PDM-176 and PDM-175 (SEQ ID NO: 352 and 353, respectively), and 1 µl PET28Ra3/38kD/38-1/Ra2A-12 DNA at 110 ng/µl. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 71 °C for 15 sec and 72 °C for 5 min and 40 sec; and finally by annealing at 72 °C for 4 min. The two sets of primers PDM-171, PDM-172, and PDM-173, PDM-174 were annealed by heating to 95 °C for 2 min and then ramping down to 25 °C slowly at 0.1 °C/sec. DPEP DNA was used to perform PCR as described above. The 38 kD fragment was digested with Eco RI (New England Biolabs) and cloned into a modified pT7ΔL2 vector which was cut with Eco 72 I (Promega) and Eco RI. The modified pT7ΔL2 construct was designed to have a MGHHHHHH amino acid coding region in frame just 5' of the Eco 72 I site. The construct was digested with Kpn 2I (Gibco, BRL) and Pst I (New England Biolabs) and the annealed sets of phosphorylated primers (PDM-171, PDM-172 and PDM-173, PDM-174) were cloned in. The DPEP PCR fragment was digested with Eco RI and Eco 72 I and cloned into this second construct which was digested with Eco 47 III (New England Biolabs) and Eco RI. Ligations were done with a ligation kit from Panvera (Madison, WI). The resulting construct was digested with NdeI (New England Biolabs) and Eco RI, and transferred to a

10084843-022502

modified pET28 vector. The fusion construct was confirmed to be correct by DNA sequencing.

Recombinant protein was prepared essentially as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-8) are provided in SEQ ID NO: 354 and 355, respectively.

One of skill in the art will appreciate that the order of the individual antigens within the fusion protein may be changed and that comparable activity would be expected provided each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the construction of fusion proteins.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

10084843.022502